



TUGAS AKHIR – SS141501

**KLASIFIKASI KELAS RESIKO PASIEN PNEUMONIA
MENGUNAKAN METODE *HYBRID* ANALISIS
DISKRIMINAN LINIER-*PARTICLE SWARM*
OPTIMIZATION (ADL-PSO) DAN *NAÏVE BAYES*
*CLASSIFICATION***

**NENI ALYA FIRDAUSANTI
NRP 1313 100 112**

**Dosen Pembimbing
Irhamah, M.Si, Ph.D
Dr. Laksmi Wulandari, dr., Sp.P (K), FCCP**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2017**



TUGAS AKHIR – SS141501

**KLASIFIKASI KELAS RISIKO PASIEN
PNEUMONIA MENGGUNAKAN METODE *HYBRID*
ANALISIS DISKRIMINAN LINIER-*PARTICLE*
SWARM OPTIMIZATION (ADL-PSO) DAN *NAÏVE*
*BAYES CLASSIFICATION***

**NENI ALYA FIRDAUSANTI
NRP 1313 100 112**

**Dosen Pembimbing
Irhamah, M.Si, Ph.D
Dr. Laksmi Wulandari, dr., Sp.P (K)., FCCP**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2017**



FINAL PROJECT – SS 141501

***CLASSIFICATION IN PNEUMONIA RISK CLASSES
USING HYBRID LINIER DISCRIMINANT
ANALYSIS-PARTICLE SWARM OPTIMIZATION
(LDA-PSO) AND NAÏVE BAYES CLASSIFICATION***

**NENI ALYA FIRDAUSANTI
NRP 1313 100 112**

**Supervisor
Irhamah, M.Si, Ph.D
Dr. Laksmi Wulandari, dr., Sp.P (K)., FCCP**

**UNDERGRADUATE PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2017**

LEMBAR PENGESAHAN

KLASIFIKASI KELAS RESIKO PASIEN PNEUMONIA MENGUNAKAN METODE *HYBRID ANALISIS* *DISKRIMINAN LINIER-PARTICLE SWARM* *OPTIMIZATION (ADI-PSO) DAN NAÏVE BAYES* *CLASSIFICATION*

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada
Program Studi Sarjana Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember

Oleh :

Neni Alya Firdausanti
NRP. 1313 100 112

Disetujui oleh Pembimbing:

Irhamah, M.Si, Ph.D

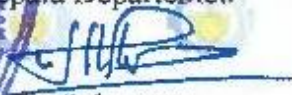
NIP. 19780406 200112 2 002

Dr. Laksmi Wulandari, dr. Sp.P (K), FCCP ()

NIP. 19680516 199901 2 001



Mengetahui,
Kepala Departemen


Dr. Suhartono

NIP. 19710929 199512 1 001

SURABAYA, JULI 2017

**KLASIFIKASI KELAS RISIKO PASIEN PNEUMONIA
MENGUNAKAN METODE *HYBRID ANALISIS
DISKRIMINAN LINIER-PARTICLE SWARM
OPTIMIZATION (ADL-PSO)* DAN *NAÏVE BAYES
CLASSIFICATION***

Nama Mahasiswa : Neni Alya Firdausanti
NRP : 1313 100 112
Departemen : Statistika
Dosen Pembimbing 1 : Irhamah, M.Si, Ph.D
Dosen Pembimbing 2 : Dr. Laksmi Wulandari, dr.,Sp.P
(K)., FCCP

Abstrak

Pneumonia merupakan penyakit yang menular melalui udara sehingga penyebaran penyakit ini tergolong sangat cepat. Oleh karena itu diagnosa yang cepat dan tepat sangat diperlukan agar dapat menentukan penanganan dan perawatan yang tepat. Beberapa penilaian kelas risiko pneumonia dikembangkan untuk mempermudah diagnosis pneumonia. Terdapat berbagai sistem skoring PSI, CURB-65, modified ATS (m-ATS) dan sebagainya untuk menentukan kelas risiko pasien pneumonia, akan tetapi tidak ada patokan pasti untuk menentukan sistem apa yang harus digunakan untuk mengelompokkan kelas risiko pneumonia. Oleh karena itu diperlukan studi klasifikasi untuk mengkaji variabel-variabel yang dapat digunakan untuk mengelompokkan kelas risiko pneumonia secara tepat. Penelitian ini menerapkan metode Analisis Diskriminan Linier dengan seleksi variabel forward selection, backward elimination dan stepwise method, Hybrid Analisis Diskriminan Linier-Particle Swarm Optimization (ADL-PSO) dan Naïve Bayes untuk mengklasifikasikan kelompok kelas risiko pneumonia berdasarkan data rekam medis pasien kemudian ketiga metode klasifikasi tersebut dibandingkan nilai akurasi. Hasil penelitian menunjukkan bahwa metode klasifikasi terbaik adalah ADL-PSO.

Kata Kunci : Analisis Diskriminan Linier, Naïve Bayes, Particle Swarm Optimization, Pneumonia, Seleksi Variabel

(Halaman ini sengaja dikosongkan)

**CLASSIFICATION IN PNEUMONIA RISK CLASSES
USING HYBRID LINIER DISCRIMINANT ANALYSIS-
PARTICLE SWARM OPTIMIZATION (LDA-PSO) AND
NAÏVE BAYES CLASSIFICATION**

Name : Neni Alya Firdausanti
NRP : 1313 100 112
Major : Statistics
Supervisor 1 : Irhamah, M.Si, Ph.D
Supervisor 2 : Dr. Laksmi Wulandari, dr., Sp.P
(K)., FCCP

Abstract

Pneumonia is a disease that is transmitted through the air so that the spread of this disease is very fast. Therefore a fast and precise diagnosis is necessary in order to determine appropriate treatment and care. Several scoring assessments of pneumonia were developed to facilitate the diagnosis of pneumonia. There are PSI scoring systems, CURB-65, modified ATS (m-ATS) and so on to determine the risk class of pneumonia patients, but there is no definite benchmark to determine what system should be used to classify the risk class of pneumonia. Therefore, a classification study is needed to assess the variables used to correctly classify the risk of pneumonia. This research applies Linear Discriminant Analysis method with selection of forward selection, backward elimination and stepwise method, Hybrid Linear Discriminant Analysis-Particle Swarm Optimization (LDA-PSO) and Naïve Bayes to classify pneumonia risk class group based on patient medical record data then all classification methods are compared the value of its accuracy. The results showed that the best classification method is LDA-PSO.

Keywords : *Linear Discriminant Analysis, Naïve Bayes, Particle Swarm Optimization, Pneumonia, Variable Selection*

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur kehadirat Allah SWT tak pernah berhenti terucap atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul **“Klasifikasi Kelas Risiko Pasien Pneumonia Menggunakan Metode Hybrid Analisis Diskriminan Linier-Particle Swarm Optimization (ADL-PSO) dan Naïve Bayes Classification”** dengan lancar.

Keberhasilan penyusunan Tugas Akhir ini tidak lepas dari banyaknya bantuan dan dukungan yang diberikan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis mengucapkan terima kasih kepada:

1. Irhamah, M.Si, Ph.D dan Dr. Laksmi Wulandari, dr.,Sp.P (K)., FCCP selaku dosen pembimbing yang telah sabar dalam memberikan bimbingan, saran, dan dukungan selama penyusunan Tugas Akhir disela-sela kesibukan yang sangat padat.
2. Ibu Santi Wulan Purnami, S.Si, M.Si dan Ibu Pratnya Paramitha O.M.Si selaku dosen penguji yang telah memberikan banyak bantuan dan masukan untuk kesempurnaan Tugas Akhir ini.
3. Bapak Hasan dan Ibu Siti Mujayanah selaku orangtua penulis yang selalu memberikan dukungan dan kasih sayang, yang menjadi motivasi bagi penulis.
4. Bapak Dr. Suhartono selaku Ketua Jurusan Statistika ITS dan Bapak Dr. Sutikno, S.Si, M.Si selaku ketua prodi S1 Statistika ITS dan dosen wali yang selalu memberi nasehat dan dukungan selama 8 semester.
5. Seluruh dosen Jurusan Statistika ITS yang telah memberikan ilmu dan pengetahuan yang tak ternilai harganya, serta segenap karyawan Jurusan Statistika ITS.

Penulis berharap hasil Tugas Akhir ini dapat bermanfaat bagi kita semua.

Surabaya, Juli 2017

Penulis

(Halaman ini sengaja dikosongkan)

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
COVER PAGE	iii
LEMBAR PENGESAHAN	v
ABSTRAK	vii
ABSTRACT	ix
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xix
 BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Perumusan Masalah	4
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	5
1.5 Batasan Penelitian	6
 BAB II TINJAUAN PUSTAKA	
2.1 Analisis Diskriminan	7
2.2 Estimasi Parameter Fungsi Diskriminan Linier..	13
2.3 <i>Particle Swarm Optimization</i> (PSO)	14
2.4 Klasifikasi Naïve Bayes	15
2.5 Pengukuran Akurasi Klasifikasi	17
2.6 Pneumonia	19
2.7 Diagnosis Pneumonia	21
 BAB III METODOLOGI PENELITIAN	
3.1 Sumber Data	25
3.2 Kerangka Konsep	25
3.3 Variabel Penelitian	25
3.4 Struktur Data Penelitian	26
3.5 Langkah Analisis	28
3.6 Diagram Alir	30

BAB IV ANALISIS DAN PEMBAHASAN

4.1	Karakteristik Data Pasien Pneumonia di RSUD Dr. Soetomo	35
4.2	Analisis Diskriminan Pasien Pneumonia	50
4.2.1	Analisis Diskriminan dengan <i>Forward Selection</i>	55
4.2.2	Analisis Diskriminan dengan <i>Backward Elimination</i>	58
4.2.3	Analisis Diskriminan dengan <i>Stepwise Method</i>	60
4.2.4	Analisis Diskriminan – PSO	61
4.3	Klasifikasi Naïve Bayes Pasien Pneumonia	70
4.4	Perbandingan Hasil Klasifikasi Analisis Diskriminan dan Naïve Bayes	74

BAB V KESIMPULAN DAN SARAN

5.1	Kesimpulan	77
5.2	Saran	78

DAFTAR PUSTAKA	79
-----------------------------	----

LAMPIRAN	83
-----------------------	----

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Proposi Penyebab Kematian pada Umur 1-4 Tahun	20
Gambar 3.1 Kriteria Penentuan Kelas Resiko Pneumonia	27
Gambar 3.2 Diagram Alir Penelitian	31
Gambar 3.3 Diagram Alir PSO.....	32
Gambar 3.4 Diagram Alir Seleksi Variabel PSI	33
Gambar 3.5 Diagram Alir Klasifikasi Naïve Bayes	34
Gambar 4.1 Karakteristik Variabel Independen	36
Gambar 4.2 Proporsi Tiap Kelas Resiko Pneumonia	40
Gambar 4.3 Perbandingan Rata-Rata Variabel Tiap Kategori	42
Gambar 4.4 Perbedaan Rata-Rata Variabel Usia Antar Kategori	43
Gambar 4.5 Perbedaan Rata-Rata Tekanan Darah Sistolik Antar Kategori	43
Gambar 4.6 Perbedaan Rata-Rata Tekanan Darah Diastolik Antar Kategori	44
Gambar 4.7 Perbedaan Rata-Rata Denyut Nadi Antar Kategori	45
Gambar 4.8 Perbedaan Rata-Rata Frekuensi Napas Antar Kategori	45
Gambar 4.9 Perbedaan Rata-Raya Suhu Tubuh Antar Kategori	46
Gambar 4.10 Perbedaan Rata-Rata Analisis Gas Darah Arteri Antar Kategori	47
Gambar 4.11 Perbedaan Rata-Rata Kadar Natrium Darah Antar Kategori	47

Gambar 4.12	Perbedaan Rata-Rata Kadar Glukosa Darah Antar Kategori	48
Gambar 4.13	Perbedaan Rata-Rata Hematokrit Antar Kategori	49
Gambar 4.14	Perbedaan Rata-Rata Kadar PO ₂ Antar Kategori	49
Gambar 4.15	Perbedaan Rata-Rata Variabel BUN Antar Kategori	50
Gambar 4.16	Representasi Inisialisasi Posisi Partikel PSO pada Iterasi Awal	62
Gambar 4.17	Representasi Posisi <i>Global Best</i> pada Iterasi Awal	64
Gambar 4.18	Representasi Posisi Partikel pada Iterasi Kedua	65

DAFTAR TABEL

	Halaman
Tabel 2.1 <i>Confusion Matrix</i>	17
Tabel 2.2 Skor Prediksi CURB-65	22
Tabel 2.3 Kelas Risiko dengan Sistem CURB-65	22
Tabel 2.4 Sistem Skor pada Pneumonia Komunitas Sistem PSI	22
Tabel 2.5 Derajat Skor Risiko dengan Sistem PSI	23
Tabel 3.1 Variabel Penelitian	25
Tabel 3.2 Struktur Data	26
Tabel 4.1 Statistika Deskriptif Variabel Independen	35
Tabel 4.2 Statistika Deskriptif Variabel Independen Tiap Kategori	41
Tabel 4.3 Rangkuman Pengujian Asumsi Data <i>Training</i> ..	51
Tabel 4.4 Transformasi Box-Cox Variabel	51
Tabel 4.5 Rangkuman Pengujian Asumsi Data <i>Training</i>	52
Tabel 4.6 Statistik Uji Perbedaan Rata-rata antar Kelompok	53
Tabel 4.7 Korelasi Antar Variabel Independen	54
Tabel 4.8 Tahapan <i>Forward Selection</i> Data Pasien Pneumonia	56
Tabel 4.9 <i>Confusion Matrix</i> Klasifikasi Diskriminan <i>Forward</i> Data <i>Testing</i>	57
Tabel 4.10 Tahapan <i>Backward Elimination</i> Data Pasien Pneumonia	59
Tabel 4.11 Tahapan <i>Stepwise Method</i> Data Pasien Pneumonia	60
Tabel 4.12 Ilustrasi Populasi Awal Partikel PSO	63

Tabel 4.13	Ilustrasi Nilai <i>Fitness</i> Setiap Partikel Populasi Awal PSO	63
Tabel 4.14	Seleksi Variabel dengan Algoritma PSO dengan <i>Fitness G-mean</i>	65
Tabel 4.15	<i>Confusion Matrix</i> Klasifikasi ADL-PSO Data Testing	66
Tabel 4.16	Seleksi Variabel dengan Algoritma PSO dengan <i>Fitness AUC</i>	67
Tabel 4.17	Hasil Seleksi Variabel	68
Tabel 4.18	Rangkuman Estimasi Parameter Fungsi Diskriminan	68
Tabel 4.19	Rata-rata dan Standar Deviasi Tiap Kategori Kelas Risiko	70
Tabel 4.20	Peluang Parsial pada Tiap Kategori Data Testing Pertama	71
Tabel 4.21	<i>Confusion Matrix</i> Klasifikasi Naive Bayes Data Testing	72
Tabel 4.22	Hasil Klasifikasi Naïve Bayes	73
Tabel 4.23	Akurasi Klasifikasi Tiap Metode	75

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Data <i>Training</i> Pasien Pneumonia	83
Lampiran 2 Data <i>Testing</i> Pasien Pneumonia	84
Lampiran 3 <i>Syntax</i> Estimasi Parameter Fungsi Diskriminan	85
Lampiran 4 <i>Syntax</i> Pembagian <i>Training</i> dan <i>Testing</i>	89
Lampiran 5 <i>Syntax</i> Seleksi Variabel <i>Forward Selection</i>	90
Lampiran 6 <i>Syntax</i> Seleksi Variabel <i>Backward Elimination</i>	93
Lampiran 7 <i>Syntax</i> Algoritma PSO	95
Lampiran 8 Sertifikat Kode Etik Penelitian di RSUD Dr. Soetomo	100

(Halaman ini sengaja dikosongkan)

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pneumonia merupakan salah satu penyakit infeksi akut pada saluran pernafasan bagian bawah yang berpengaruh pada paru-paru. Pengertian lain dari pneumonia adalah peradangan paru yang disebabkan oleh mikroorganisme seperti bakteri, virus, jamur dan parasit (PDPI, 2003). Ketika seseorang terkena pneumonia, alveoli dipenuhi nanah dan cairan, sehingga menyebabkan bernafas menjadi menyakitkan dan oksigen yang masuk ke paru-paru hanya sedikit. Pneumonia merupakan penyebab kematian terbesar baik di negara-negara berkembang maupun di negara-negara maju seperti Amerika Serikat, Kanada dan negara-negara Eropa (Setyoningrum, 2006). Kematian akibat pneumonia seringkali disebabkan oleh infeksi kuman, seperti *Haemophilus Influenza* atau *Streptococcus aureus*, *Klebsiella Sp*, *Pseudomonas Sp*, dan virus, seperti *virus Influenza*. Pneumonia dapat terjadi sepanjang tahun dan dapat melanda semua usia. Manifestasi klinik menjadi sangat berat pada pasien dengan usia sangat muda atau usia di bawah lima tahun, manula, serta pada pasien dengan kondisi kritis (Kemenkes RI, 2010).

Data *SEAMIC Health Statistics* tahun 2001, menyebutkan bahwa influenza dan pneumonia merupakan urutan ke-enam penyebab kematian di Indonesia, nomor sembilan di Brunei, nomor tujuh di Malaysia, nomor tiga di Singapura, nomor enam di Thailand, dan nomor tiga di Vietnam. Salah satu kelompok berisiko tinggi terkena pneumonia adalah usia 65 tahun atau lebih (American Lung Association, 2015). Usia lanjut yang terkena pneumonia memiliki derajat keparahan yang tinggi, bahkan dapat menyebabkan kematian. *World Health Organization* (WHO) tahun 2005 memperkirakan kematian akibat pneumonia pada usia lanjut berkisar 167 per 100.000 penduduk, dimana sekitar 70% terjadi di negara-negara berkembang terutama Afrika dan Asia Tenggara. Data Risdas 2013, *period prevalence* pneumonia di Indonesia

yang tinggi terjadi pada kelompok umur 1-4 tahun yaitu sebesar 25,8%, kemudian meningkat pada kelompok umur 45-54 tahun yaitu sebesar 12,8% dan terus meningkat pada kelompok umur berikutnya. Beberapa rumah sakit besar di Indonesia mencatat bahwa kematian yang disebabkan pneumonia masih cukup tinggi, salah satunya adalah RSUD Dr. Soetomo. RSUD Dr. Soetomo merupakan rumah sakit terbesar dan terlengkap di Jawa Timur, sehingga rumah sakit ini menjadi rujukan beberapa kasus tertentu yang tidak bisa ditangani oleh rumah sakit lainnya, baik di wilayah Jawa Timur maupun wilayah Indonesia bagian timur lainnya. Data yang didapatkan RSUD Dr. Soetomo menunjukkan bahwa terdapat sekitar 180 kasus pneumonia komunitas dengan angka kematian antara 20% hingga 30% (PDPI, 2003).

Penyebab pneumonia sangat sulit ditemukan dan memerlukan waktu beberapa hari untuk mendapatkan hasilnya, sedangkan apabila penderita pneumonia tidak segera diobati akan menyebabkan kematian. Di Amerika, walaupun menggunakan cara invasif penyebab pneumonia hanya dapat ditemukan sekitar 50%. Selain itu dikarenakan penularan pneumonia melalui udara, penyakit ini termasuk dalam penyakit yang penyebarannya sangat cepat. Deteksi dini pneumonia tidak mudah karena gejala penyakit ini hampir sama dengan batuk. Selain deteksi dini, penilaian kelas risiko pneumonia merupakan komponen penting dalam tatalaksana Pneumonia Komunitas. Penilaian kelas risiko ini kemudian digunakan untuk menentukan penanganan yang dilakukan pada pasien pneumonia. Hal ini membuat munculnya berbagai sistem skor PSI, CURB-65, modified ATS (m-ATS) dan sebagainya untuk menentukan kelas risiko. Beberapa metode dalam pengelompokan kelas risiko yang lazim digunakan adalah pengelompokan berdasarkan sistem skor PSI (*Pneumonia Severity Index*) yang dikembangkan oleh *Pneumonia Patient Outcome Research Team* (PORT) dan sistem skor CURB-65 yang direkomendasikan oleh *British Thoracic Society* (BTS). Kedua sistem skor tersebut memiliki performa yang sama dalam mengelompokkan risiko pneumonia. Akan tetapi penggunaan

skor tersebut belum teruji pada jumlah sampel yang besar. Oleh karena itu, diperlukan analisis mengenai klasifikasi kelas risiko pneumonia untuk mengkaji variabel-variabel yang dapat digunakan untuk mengelompokkan kelas risiko pneumonia secara tepat sehingga dapat mempercepat proses penanganan atau penentuan pengobatan yang tepat bagi penderita. Penelitian sebelumnya mengenai penyakit pneumonia sudah banyak dilakukan. Noviana dkk (2013) melakukan penelitian mengenai pemodelan risiko penyakit pneumonia pada balita di Jawa Timur menggunakan regresi logistik biner stratifikasi untuk mendapatkan model faktor-faktor yang mempengaruhi penyakit pneumonia. Metode *Geographically Wighted Logistic Regression* juga pernah diterapkan oleh Fatimah dkk (2015) untuk memodelkan risiko penyakit pneumonia pada balita di provinsi Jawa Timur.

Beberapa metode klasifikasi dalam ilmu statistik dapat digunakan untuk menentukan kelas risiko pasien pneumonia, salah satunya adalah analisis diskriminan. Analisis diskriminan merupakan teknik multivariat yang berkaitan dengan mengelompokkan suatu objek atau observasi baru terhadap kelompok yang telah didefinisikan sebelumnya (Johnson & Wichern, 2007). Tujuan utama analisis diskriminan adalah untuk mengetahui perbedaan antar kelompok (Hair, Black, & Babin, 2010). Variabel independen dalam analisis diskriminan berupa variabel kontinyu dan variabel dependen berupa variabel kategori.

Seiring berkembangnya ilmu pengetahuan, banyak algoritma optimasi yang dapat dikombinasikan dengan analisis diskriminan dengan tujuan untuk meningkatkan akurasi klasifikasi. Penelitian yang dilakukan Irhamah dan Kurnianto (2016) mengenai seleksi variabel dan estimasi parameter menggunakan metode *hybrid* analisis diskriminan *multi-class* dan algoritma genetika untuk klasifikasi menunjukkan bahwa dengan menambahkan algoritma genetika dalam seleksi variabel dapat meningkatkan nilai akurasi dalam klasifikasi. Dalam penelitian ini algoritma optimasi yang ditambahkan untuk seleksi variabel adalah *Particle Swarm Optimization* (PSO). Penelitian yang dilakukan oleh Lin dan Chen

(2009) menyebutkan bahwa analisis diskriminan linier menggunakan optimasi seleksi variabel PSO menghasilkan akurasi yang lebih tinggi dibandingkan seleksi variabel dengan metode *forward*, *backward* dan PCA. Penelitian lain dilakukan Bai (2010) mengenai algoritma *Particle Swarm Optimization* menyebutkan bahwa optimasi menggunakan Algoritma PSO memiliki kelebihan, diantaranya tidak ada *overlapping* pada perhitungan, waktu pencarian untuk hasil optimal sangat cepat, dan perhitungan PSO sangat sederhana dibandingkan metode klasifikasi lainnya.

Pada penelitian ini selain melakukan klasifikasi menggunakan analisis diskriminan juga dilakukan klasifikasi menggunakan *Naïve Bayes Classification*. Klasifikasi menggunakan Naïve Bayes memiliki beberapa kelebihan dibandingkan metode lain, diantaranya Naïve Bayes memiliki desain proses yang sederhana karena tidak ada asumsi tertentu yang harus dipenuhi, mudah untuk mengolah data yang memiliki *range* variabel yang luas dan secara komputasi metode Naïve Bayes lebih cepat dalam membuat keputusan (Ashari, Paryudi, & Tjoa, 2013). Wicaksana (2015) melakukan penelitian untuk membandingkan Algoritma *K-Nearest Neighbors* dan Naïve Bayes pada data *wisconsin diagnosis breast cancer* dan dihasilkan kesimpulan bahwa algoritma Naïve Bayes mempunyai akurasi yang lebih tinggi dibandingkan dengan algoritma *K-Nearest Neighbors*, yaitu dengan akurasi sebesar 97,7% untuk Naïve Bayes dan 95,8% untuk KNN.

Berdasarkan uraian latar belakang yang telah disebutkan sebelumnya, maka akan digunakan algoritma *Particle Swarm Optimization* (PSO) untuk seleksi variabel pada Analisis Diskriminan Linier dan metode *Naïve Bayes Classification* untuk mengklasifikasikan kelas risiko pasien pneumonia di RSUD Dr. Soetomo.

1.2 Perumusan Masalah

Jumlah kematian akibat pneumonia menunjukkan angka yang cukup tinggi baik di negara maju seperti Amerika Serikat maupun di negara-negara berkembang termasuk Indonesia. Pneumonia memerlukan penanganan dan tindakan yang terarah agar kematian

dapat dicegah. Diagnosa pneumonia yang tepat akan menentukan jenis pengobatan yang diterima pasien. Berdasarkan uraian tersebut permasalahan yang akan dibahas dalam penelitian ini adalah bagaimana proses klasifikasi kelas risiko pasien pneumonia menggunakan metode Analisis Diskriminan Linier, *Hybrid Analisis Diskriminan Linier- Particle Swarm Optimization* (ADL-PSO), dan Naïve Bayes serta perbandingan akurasi dari ketiga metode tersebut.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan di atas, tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Mendapatkan karakteristik pasien penumonia RSUD Dr. Soetomo.
2. Mengetahui klasifikasi kelas risiko pasien pneumonia dan nilai akurasi dengan menggunakan metode Analisis Diskriminan Linier, *Hybrid Analisis Diskriminan Linier-Particle Swarm Optimization* (ADL-PSO) dan Naïve Bayes.
3. Membandingkan ketepatan klasifikasi antara metode Analisis Diskriminan Linier, *Hybrid Analisis Diskriminan Linier-Particle Swarm Optimization* (ADL-PSO) dan Naïve Bayes

1.4 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut.

1. Dalam bidang kedokteran, penelitian ini diharapkan dapat memberikan tambahan informasi mengenai klasifikasi penentuan kelompok kelas risiko pasien pneumonia sehingga dapat menentukan perawatan dan penanganan yang sesuai.
2. Bagi dunia pendidikan, penelitian ini diharapkan dapat dijadikan sebagai referensi untuk melakukan penelitian selanjutnya terutama bagi penelitian yang berhubungan dengan *hybrid* analisis diskriminan dan Naïve Bayes.

1.5 Batasan Penelitian

Batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Data pasien yang digunakan adalah pasien pneumonia sebagai diagnosis utama.
2. Variabel independen yang digunakan dalam penelitian ini adalah variabel kontinyu.
3. Kelas resiko yang digunakan dalam penelitian ini adalah kelas resiko II, III, IV dan V.

BAB II

TINJAUAN PUSTAKA

Tinjauan pustaka yang digunakan pada penelitian ini meliputi analisis diskriminan, *Particle Swarm Optimization* (PSO), dan pneumonia yang akan dijelaskan sebagai berikut.

2.1 Analisis Diskriminan

Analisis diskriminan merupakan salah satu metode multivariat yang bertujuan untuk memisahkan beberapa kelompok data yang sudah terkelompokan dengan cara membentuk fungsi diskriminan (Johnson & Wichern, 2007). Analisis diskriminan termasuk dalam analisis multivariat dengan metode dependensi karena variabel-variabelnya tidak saling bergantung satu dengan yang lain. Jika variabel dependen terdiri dari dua kelompok maka disebut *Two-Group Discriminant Analysis*, sedangkan jika terdapat lebih dari dua kelompok maka disebut *Multiple-class Discriminant Analysis*. Analisis diskriminan dapat digunakan sebagai metode klasifikasi untuk mengelompokkan individu ke dalam satu dari dua kelompok atau lebih.

Persamaan linier fungsi diskriminan adalah sebagai berikut (Hair, Black, & Babin, 2010).

$$\hat{Z}_{jk} = \alpha + w_1x_{1k} + w_2x_{2k} + \dots + w_px_{pk}, \quad (2.1)$$
$$k = 1, 2, \dots, n$$

Keterangan :

\hat{Z}_{jk} : Skor diskriminan dari fungsi diskriminan ke- j pengamatan ke- k , $j = g - 1$

α : Konstanta

w_i : Koefisien fungsi diskriminan untuk variabel independen ke- i , $i = 1, 2, \dots, p$

x_{ik} : Variabel independen ke- i pengamatan ke- k .

Suatu fungsi diskriminan layak dibentuk apabila terdapat perbedaan nilai rata-rata di antara kelompok-kelompok yang ada. Perbedaan antar kelompok pada analisis diskriminan dapat

diketahui dengan menghitung nilai *Wilk's Lambda* (Λ). Persamaan *Wilk's Lambda* adalah sebagai berikut (Johnson & Wichern, 2007).

$$\Lambda = \frac{|W|}{|W + B|} \quad (2.2)$$

Dimana

$$W = \left[\sum_{l=1}^g (n_l - 1) \right] S_{pooled}$$

$$S_{pooled} = \frac{1}{\sum_{l=1}^g (n_l - 1)} \{ (n_1 - 1)S_1 + (n_2 - 1)S_2 + \dots + (n_g - 1)S_g \}$$

$$B = \sum_{l=1}^g n_l (\mathbf{x}_l - \bar{\mathbf{x}})(\mathbf{x}_l - \bar{\mathbf{x}})'$$

Keterangan:

W : Matriks keragaman *within group*

B : Matriks keragaman *between group*

S_{pooled} : Matriks varian-kovarian dalam kelompok gabungan

Wilk's Lambda bernilai antara nol sampai satu. Apabila nilai *Wilk's Lambda* mendekati nol maka terdapat perbedaan rata-rata antar kelompok. Sebaliknya, apabila nilai *Wilk's Lambda* mendekati satu maka tidak terdapat perbedaan rata-rata antar kelompok.

Beberapa asumsi yang harus dipenuhi dalam analisis diskriminan adalah sebagai berikut (Santoso, 2010).

1) Normal Multivariat

Variabel independen yang digunakan dalam analisis diskriminan harus berdistribusi normal multivariat. Distribusi normal multivariat merupakan perluasan dari distribusi normal univariat dengan jumlah variabel $p \geq 2$. *Probability density function* untuk distribusi normal multivariat adalah sebagai berikut.

$$f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{p/2}} e^{-\frac{1}{2}(\mathbf{X}-\mu)' \Sigma^{-1}(\mathbf{X}-\mu)} \quad (2.3)$$

Salah satu metode yang digunakan untuk uji normalitas pada data multivariat adalah uji Mardia.

Hipotesis untuk uji Normal Multivariat adalah

H_0 : Data memenuhi asumsi Normal Multivariat

H_1 : Data tidak memenuhi asumsi Normal Multivariat

Uji normalitas multivariat Mardia didasarkan pada ekstensi multivariat dari *skewness* ($\hat{\gamma}_{1,p}$) dan kurtosis ($\hat{\gamma}_{2,p}$) yang diukur menggunakan rumus sebagai berikut.

$$\hat{\gamma}_{1,p} = \frac{1}{n^2} \sum_{k=1}^n \sum_{r=1}^n m_{kr}^3 \quad (2.4)$$

$$\hat{\gamma}_{2,p} = \frac{1}{n^2} \sum_{k=1}^n m_{kk}^2 \quad (2.5)$$

dimana,

$$m_{kr} = (\mathbf{x}_k - \bar{\mathbf{x}})' S^{-1} (\mathbf{x}_r - \bar{\mathbf{x}})$$

Statistik uji untuk *skewness* adalah $(\frac{n}{6})\hat{\gamma}_{1,p}$ didekatkan dengan distribusi *chi square* dengan derajat bebas $\frac{p(p+1)(p+2)}{6}$. Statistik uji untuk kurtosis adalah $\hat{\gamma}_{2,p}$ didekatkan dengan distribusi normal dengan rata-rata $(p(p+2))$ dan varian $\frac{8p(p+2)}{n}$.

Pada ukuran sampel kecil ($n < 20$), rawan terjadi error tipe 1, sehingga Mardia (1974) memperkenalkan koreksi untuk statistik uji *skewness* apabila jumlah sampel kurang dari 20. Statistik uji *skewness* terkoreksi pada jumlah sampel data yang kecil adalah $(\frac{nk}{6})\hat{\gamma}_{1,p}$, dimana $k = \frac{(p+1)(n+1)(n+3)}{n(n+1)(p+1)-6}$. Statistik uji tersebut juga berdistribusi *chi square* dengan derajat bebas $\frac{p(p+1)(p+2)}{6}$. (Korkmaz, Goksuluk, & Zararsiz, 2014). Dikatakan normal multivariat apabila nilai *p value* baik pada *skewness* maupun pada kurtosis lebih besar dari 0,05.

Dalam penelitian ini apabila data tidak memenuhi asumsi normal multivariat maka akan dilakukan transformasi pada data. Transformasi yang digunakan adalah transformasi Box-Cox. Transformasi Box Cox merupakan transformasi pangkat pada suatu variabel. Box Cox mempertimbangkan kelas transformasi berparameter tunggal, yaitu λ yang dipangkatkan pada variabel x , sehingga transformasinya menjadi x^λ . Box dan Cox (1964)

mendefinisikan transformasi pangkat λ sebagai berikut (Velez, Correa, & Ramos, 2015).

$$\mathbf{x}_i^{(\lambda)} = \begin{cases} \lambda^{-1}(\mathbf{x}_i^\lambda - 1) & , \text{ jika } \lambda \neq 0 \\ \text{Log}(\mathbf{x}_i) & , \text{ jika } \lambda = 0 \end{cases} \quad (2.6)$$

Dimana $\mathbf{x}^{(\lambda)}$ merupakan data yang telah ditransformasi.

2) Matriks Varian-Kovarian Homogen

Pengujian homogenitas dilakukan untuk menguji apakah matriks varian varian-kovarian antar populasi bersifat homogen atau tidak. Salah satu metode yang digunakan adalah *Box's-M test* dengan hipotesis sebagai berikut (Johnson & Wichern, 2007).

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma$$

H_1 : Minimal ada satu matriks varian-kovarian yang berbeda

Dengan g merupakan banyaknya kategori.

Statistik uji yang digunakan adalah sebagai berikut.

$$C = \frac{M}{b} \quad (2.7)$$

Dimana

$$M = \left[\sum_{l=1}^g (n_l - 1) \right] \ln |S_{pooled}| - \sum_{l=1}^g [(n_l - 1) \ln |S_l|]$$

$$b = \frac{f_1}{1 - A_1 - \frac{f_1}{f_2}}$$

$$f_1 = \frac{1}{2}(g-1)p(p+1) \quad f_2 = \frac{f_1 + 2}{A_2 - A_1^2}$$

$$A_1 = \frac{2p^2 + 3p - 1}{6(g-1)(p+1)} \left(\sum_{l=1}^g \frac{1}{n_l} - \frac{1}{N} \right)$$

$$A_2 = \frac{(p-1)(p+2)}{6(g-1)} \left(\sum_{l=1}^g \frac{1}{n_l^2} - \frac{1}{N^2} \right)$$

Keterangan:

p : Jumlah variabel independen.

S_l : Matriks varian-kovarian kelompok ke- l

n_l : Jumlah observasi kelompok ke- l

N : Total Pengamatan

f_1 : Derajat bebas 1

f_2 : Derajat bebas 2

Tolak H_0 apabila $C > F_{(f_1, f_2, \alpha)}$ dan disimpulkan bahwa terdapat perbedaan pada matriks varian-kovarian.

3) Tidak Terdapat Kasus Multikolinieritas

Multikolinieritas adalah keadaan dimana terdapat korelasi antar variabel prediktor. Multikolinieritas dapat dideteksi dengan menguji korelasi antar variabel independen (Ghozali, 2011). Nilai signifikansi korelasi antar variabel dapat diuji menggunakan hipotesis sebagai berikut.

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Statistik uji yang digunakan adalah sebagai berikut.

$$t_{hitung} = \hat{\rho}_{xy} \sqrt{\frac{n-2}{1-\hat{\rho}_{xy}^2}} \quad (2.8)$$

Dimana

$$\hat{\rho}_{xy} = \frac{\sum_{k=1}^n (\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{y}_k - \bar{\mathbf{y}})}{\sqrt{\sum_{k=1}^n (\mathbf{x}_k - \bar{\mathbf{x}})^2} \sqrt{\sum_{k=1}^n (\mathbf{y}_k - \bar{\mathbf{y}})^2}}$$

Keterangan

$\hat{\rho}_{xy}$: Korelasi antar variabel x dengan variabel y

n : banyaknya observasi

Tolak H_0 jika $|t_{hitung}| > t_{(n-2, \frac{\alpha}{2})}$ atau $p_{value} < 0,05$ sehingga disimpulkan terdapat korelasi antar variabel independen.

Deteksi multikolinieritas juga dapat dilakukan dengan menggunakan nilai VIF (*Variance Inflation Factor*). Nilai VIF dapat dihitung dengan menggunakan rumus sebagai berikut (Hocking, 2003).

$$VIF = \frac{1}{(1 - R^2_l)} \quad (2.9)$$

Semakin besar koefisien korelasi antar variabel maka nilai VIF akan semakin besar. Jika nilai VIF lebih dari 10, maka

dikatakan bahwa peubah X_l memiliki hubungan yang erat dengan peubah X lainnya atau dengan kata lain terdapat kasus multikolinieritas.

Dalam analisis diskriminan diperlukan seleksi variabel. Seleksi variabel dilakukan karena terdapat kemungkinan tidak semua variabel berpengaruh dalam analisis. Seleksi variabel pada analisis diskriminan dilakukan menggunakan nilai *Wilks' Λ* sebagai kriteria penilaian. Metode seleksi variabel yang biasa digunakan antara lain *stepwise method*, *forward selection*, dan *backward elimination*. Gabungan dari *forward selection* dan *backward elimination* adalah metode *stepwise method*. Metode seleksi variabel *forward selection* dilakukan dengan memasukkan variabel independen satu per satu yang memiliki nilai F_{hitung} parsial yang paling besar berdasarkan nilai *Wilks' Λ* dan $p_{value} < \alpha$. Persamaan *Wilks' Λ* dan F_{hitung} adalah sebagai berikut .

$$\Lambda(x_r | x_1, x_2, \dots, x_p) = \frac{\Lambda(x_1, x_2, \dots, x_r, x_p)}{\Lambda(x_1)} \quad (2.10)$$

$$F_{hitung} = \left(\frac{1 - \Lambda(x_r | x_1, x_2, \dots, x_p)}{\Lambda(x_r | x_1, x_2, \dots, x_p)} \right) \left(\frac{n - g - p}{g - 1} \right) \quad (2.11)$$

Backward elimination merupakan seleksi variabel dengan mengeliminasi satu persatu variabel yang memiliki nilai F_{hitung} parsial yang rendah berdasarkan nilai *Wilks' Λ* dan $p_{value} \geq \alpha$. Persamaan *Wilks' Λ* dan F_{hitung} dalam metode *backward elimination* adalah sebagai berikut.

$$\Lambda(x_r | x_1, x_2, \dots, x_p) = \frac{\Lambda(x_1, x_2, \dots, x_r, x_p)}{\Lambda(x_1)} \quad (2.12)$$

$$F_{hitung} = \left(\frac{1 - \Lambda(x_r | x_1, x_2, \dots, x_p)}{\Lambda(x_r | x_1, x_2, \dots, x_p)} \right) \left(\frac{n - g - p + 1}{g - 1} \right) \quad (2.13)$$

Dalam metode *stepwise method* apabila variabel yang dimasukkan memiliki nilai *Wilks' Λ* , maka variabel tersebut akan dieeliminasi menggunakan *backward elimination*.

2.2 Estimasi Parameter Fungsi Diskriminan Linier

Pada g kelompok populasi dengan ukuran masing-masing n_l , $l = 1, 2, \dots, g$, vektor variabel independen kelompok populasi ke- l dengan $\mathbf{X}_l = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$ dan baris ke- k adalah \mathbf{x}_{lk} maka vektor variabel rata-rata populasi ke- l dapat dinyatakan sebagai berikut.

$$\bar{\mathbf{x}}_l = \frac{1}{n_l} \sum_{k=1}^{n_l} \mathbf{X}_{lk} \quad (2.14)$$

dan vektor rata-rata populasi adalah sebagai berikut.

$$\bar{\mathbf{x}} = \frac{\sum_{l=1}^g n_l \bar{\mathbf{x}}_l}{\sum_{l=1}^g n_l} = \frac{\sum_{l=1}^g \sum_{k=1}^{n_l} \mathbf{x}_{lk}}{\sum_{l=1}^g n_l} \quad (2.15)$$

Fisher mengembangkan konsep dasar dalam pengembangan analisis diskriminan dengan memaksimumkan nilai rasio matriks keragaman *between group* terhadap matriks keragaman *within group*. Nilai rasio dihitung menggunakan rumus sebagai berikut (Johnson & Wichern, 2007).

$$\frac{\hat{\mathbf{a}}' \mathbf{B} \hat{\mathbf{a}}}{\hat{\mathbf{a}}' \mathbf{W} \hat{\mathbf{a}}} = \frac{\hat{\mathbf{a}}' (\sum_{l=1}^g (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})') \hat{\mathbf{a}}}{\hat{\mathbf{a}}' (\sum_{l=1}^g \sum_{k=1}^{n_l} (\mathbf{x}_{lk} - \bar{\mathbf{x}}_l)(\mathbf{x}_{lk} - \bar{\mathbf{x}}_l)') \hat{\mathbf{a}}} \quad (2.16)$$

Keterangan :

- $\hat{\mathbf{a}}$: Vektor koefisien fungsi diskriminan
- \mathbf{B} : Matriks keragaman *between group*
- \mathbf{W} : Matriks keragaman *within group*
- $\bar{\mathbf{x}}_l$: Vektor rata-rata kategori ke- l
- $\bar{\mathbf{x}}$: Vektor rata-rata pengamatan
- \mathbf{x}_{lk} : Vektor variabel bebas pada pengamatan ke- k dan kategori ke- l

Setelah didapatkan fungsi diskriminan, langkah selanjutnya adalah melakukan validasi fungsi diskriminan dengan menggunakan matriks klasifikasi untuk mendapatkan nilai ketepatan klasifikasi. Tahapan awal yang dapat dilakukan untuk mendapatkan nilai ketepatan klasifikasi adalah mengestimasi parameter, kemudian melakukan klasifikasi. Koefisien parameter fungsi diskriminan yang digunakan pada penelitian ini merupakan

koefisien *canonical* yang tidak terstandarisasi yang diperoleh dari rumus berikut ini.

$$\mathbf{w} = \sqrt{n - p} \hat{\mathbf{a}} \quad (2.17)$$

$$\boldsymbol{\alpha} = -\bar{\mathbf{x}}\mathbf{w} \quad (2.18)$$

Keterangan:

- \mathbf{w} = vektor koefisien fungsi diskriminan tidak terstandarisasi
 n = jumlah pengamatan
 p = jumlah kategori
 $\boldsymbol{\alpha}$ = konstanta
 $\hat{\mathbf{a}}$ = nilai *eigen vector* dari $\mathbf{W}^{-1}\mathbf{B}$

Untuk mendapatkan nilai ketepatan klasifikasi, Fisher mengelompokkan $\tilde{\mathbf{x}}$ kedalam kelompok ke- l dengan menggunakan aturan sebagai berikut.

$$\sum_{j=1}^h (\hat{Z}_j - \bar{Z}_{lj})^2 = \sum_{j=1}^h (\hat{\mathbf{a}}_j(\mathbf{x} - \bar{\mathbf{x}}_l))^2 \leq \sum_{j=1}^h (\hat{\mathbf{a}}_j(\mathbf{x} - \bar{\mathbf{x}}_s))^2, s \neq l \quad (2.19)$$

Dimana $\bar{Z}_{lj} = \hat{\mathbf{a}}_j'$

Keterangan:

$$h \leq \min(g - 1, p)$$

2.3 Particle Swarm Optimization (PSO)

Particle Swarm Optimization merupakan metode optimasi yang dikenalkan oleh Kennedy dan Eberhart pada tahun 1995 (Kennedy & Eberhart, 1995). PSO dikembangkan berdasarkan perilaku gerakan kawanan burung dan ikan dalam mencari sumber makanan. Burung akan menyebar atau pergi bersama-sama untuk mencari sumber makanan yang optimum. Arah pergerakan burung menuju sumber makanan optimum dipengaruhi oleh pergerakan yang sedang terjadi, sumber makanan terbaik yang pernah dilalui, dan sumber makanan terbaik yang pernah ditemukan oleh salah satu burung dalam kelompok. Dalam algoritma dasar PSO, segerombolan m partikel yang bergerak dipengaruhi oleh 1) inersia, 2) posisi lokal yang terbaik (*local best position*), dan 3) posisi global yang terbaik (*global best position*). Kecepatan dan

posisi setiap partikel ke- q berubah menurut persamaan berikut (Shi & Eberhart, 1998).

$$v_q^{t+1} = v_q^t + c_1 r_1^t (pbest_q^t - x_q^t) + c_2 r_2^t (gbest^t - x_q^t) \quad (2.20)$$

Dimana,

$$x_q^{t+1} = x_q^t + v_q^{t+1} \quad (2.21)$$

$t = 1, 2, \dots, T - 1$ dan $q = 1, 2, \dots, m$

Pada persamaan di atas v_q^t dan x_q^t menunjukkan kecepatan dan posisi partikel ke- q pada waktu ke- t . $pbest_q^t$ merupakan nilai individu ke- q pada posisi yang optimum (*local best position*). $gbest^t$ adalah nilai posisi terbaik dari seluruh gerombolan (*global best position*). c_1 dan c_2 merupakan suatu koefisien yang mengatur kecepatan partikel sehingga solusi yang didapatkan tidak akan menjadi optimum lokal, nilai c_1 dan c_2 yang biasa digunakan adalah 2. r_1 dan r_2 adalah random vektor berdistribusi uniform dengan nilai 0-1. Untuk menghindari partikel terlalu jauh dari ruang pencarian, kecepatan partikel dibuat pada setiap arahnya terbatas antara $-v_{dmax}$ dan v_{dmax} . Jika jumlah v_{dmax} terlalu besar, solusinya akan jauh dari yang terbaik, jika jumlah v_{dmax} terlalu kecil, solusi akan jadi optimum lokal.

Penambahan bobot inersia pada PSO pertamakali dikemukakan oleh Shi dan Eberhart pada tahun 1988 (Bai, 2010). Penambahan bobot inersia pada PSO biasa disebut *Global Best PSO* (GBPSO). GBPSO merupakan varian PSO yang paling dasar (Schoene, 2011). Iterasi kecepatan pada GBPSO ditentukan menggunakan rumus sebagai berikut (Kennedy & Eberhart, 1995).

$$v_q^{t+1} = \omega v_q^t + c_1 r_1^t (pbest_q^t - x_q^t) + c_2 r_2^t (gbest^t - x_q^t) \quad (2.22)$$

$t = 1, 2, \dots, T - 1$ dan $q = 1, 2, \dots, m$

ω merupakan bobot inersia untuk kecepatan partikel ke- q pada waktu ke- t . Bobot inersia digunakan untuk mempercepat konvergensi dan meningkatkan fungsi PSO (Bai, 2010)

2.4 Klasifikasi Naïve Bayes

Diketahui A_1, A_2, \dots, A_p merupakan kejadian saling asing di dalam ruang sampel Ω sehingga $\sum_{i=1}^p A_i = \Omega$. Kejadian A_1, A_2, \dots, A_p merupakan partisi dalam Ω . B merupakan suatu kejadian acak dan

peristiwa $A_1 \cap B, A_2 \cap B, \dots, A_p \cap B$ membentuk partisi dalam B (Gorunescu, 2011).

$$P(B) = \sum_{i=1}^p P(A_i \cap B) \quad (2.23)$$

Jika $P(A_i) > 0$ pada $i = 1, 2, \dots, p$ maka $P(A_i \cap B) = P(B|A_i)P(A_i)$.

$$P(B) = \sum_{i=1}^p P(B|A_i)P(A_i) \quad (2.24)$$

Kejadian acak B dan A_1, A_2, \dots, A_p merupakan partisi dari ruang sampel Ω . Apabila $P(B) > 0$ dan $P(A_i) > 0$ untuk $i = 1, 2, \dots, p$, maka:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^p P(B|A_i)P(A_i)} \quad (2.25)$$

$P(A_i|B)$ disebut *posterior probability* karena nilainya bergantung pada nilai B , $P(A_i)$ merupakan *prior probability* karena nilainya tidak bergantung pada B , $P(B|A_i)$ merupakan fungsi *likelihood* dan $P(B)$ merupakan keterangan (*evidence*).

Klasifikasi Naïve Bayes merupakan klasifikasi berdasarkan teorema Bayes dan digunakan untuk menghitung probabilitas tiap kelas dengan asumsi tidak ada hubungan antar kelas (independen). Diberikan $\{x_1, x_2, \dots, x_p\}$ merupakan atribut yang digunakan untuk menentukan kelas y . Perhitungan *prosterior probability* untuk setiap kelas y_l menggunakan teorema Bayes adalah sebagai berikut (Gorunescu, 2011).

$$P(y_l|x_1, x_2, \dots, x_p) = \frac{P(x_1, x_2, \dots, x_p|y_l) \cdot P(y_l)}{P(x_1, x_2, \dots, x_p)} \quad (2.26)$$

Kelas yang dipilih adalah kelas yang memaksimalkan nilai $P(y_l|x_1, x_2, \dots, x_p)$ atau memaksimalkan $P(x_1, x_2, \dots, x_p|y_l) \cdot P(y_l)$. Berdasarkan persamaan di atas maka diperlukan perhitungan $P(x_1, x_2, \dots, x_p|y_l)$. Setiap atribut diasumsikan saling bebas untuk kelas y .

$$P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p | \mathbf{y}) \quad (2.27)$$

$$= P(\mathbf{x}_1 | \mathbf{y}) \cdot P(\mathbf{x}_2 | \mathbf{y}) \dots P(\mathbf{x}_p | \mathbf{y})$$

Jika terdapat atribut yang bersifat kuantitatif atau kontinyu, maka $P(\mathbf{x}_i | \mathbf{y}_l)$ dihitung menggunakan pendekatan distribusi normal.

$$P(\mathbf{x}_i | \mathbf{y}_l) = \frac{1}{\sigma_{il} \sqrt{2\pi}} \exp\left(-\frac{(\mathbf{x}_i - \mu_{il})^2}{2\sigma_{il}^2}\right) \quad (2.28)$$

Estimasi peluang $P(\mathbf{x}_i | \mathbf{y}_l)$ dapat dihitung untuk setiap atribut \mathbf{x}_i dan kelas \mathbf{y}_l sehingga data baru akan dapat diklasifikasikan ke dalam kelas \mathbf{y}_l apabila peluang yang didapatkan paling besar diantara yang lainnya.

Dalam Naïve Bayes digunakan *Hypothesis Maximum A Posterior* (HMAP) untuk memaksimalkan nilai probabilitas masing-masing kelas dengan menggunakan rumus sebagai berikut.

$$H_{MAP} = \arg \max \frac{P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p | \mathbf{y}_l) \cdot P(\mathbf{y}_l)}{P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)} \quad (2.29)$$

2.5 Pengukuran Akurasi Klasifikasi

Nilai akurasi hasil klasifikasi dapat dihitung menggunakan *confusion matrix*. *Confusion matrix* adalah alat yang digunakan untuk menganalisis seberapa baik *classifier* mengenali data dari kelas yang berbeda.

Tabel 2.1 *Confusion Matrix*

Aktual	Prediksi	
	$\hat{\pi}_1$	$\hat{\pi}_2$
$\hat{\pi}_1$	n_{11}	n_{12}
$\hat{\pi}_2$	n_{21}	n_{22}

Keterangan:

n_{11} : jumlah prediksi $\hat{\pi}_1$ yang tepat diklasifikasikan ke π_1

n_{22} : jumlah prediksi $\hat{\pi}_2$ yang tepat diklasifikasikan ke π_2

n_{21} : jumlah prediksi $\hat{\pi}_1$ yang salah diklasifikasikan ke π_2

n_{12} : jumlah prediksi $\hat{\pi}_2$ yang salah diklasifikasikan ke π_1

Dengan menggunakan *confusion matrix*, beberapa pengukuran performa klasifikasi dapat dilakukan, yaitu akurasi, sensitivitas, spesifitas, dan presisi.

$$akurasi = \frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}} \quad (2.30)$$

$$sensitivitas = \frac{n_{11}}{n_{11} + n_{12}} \quad (2.31)$$

$$spesifitas = \frac{n_{22}}{n_{21} + n_{22}} \quad (2.32)$$

Akurasi menunjukkan efektivitas *classifier* secara menyeluruh. Semakin besar nilai akurasi maka semakin baik kinerja *classifier*. Sensitivitas digunakan untuk mengukur efektivitas sebuah *classifier* untuk mengidentifikasi kelas positif, sedangkan spesifitas mengukur efektivitas *classifier* dalam mengidentifikasi kelas negatif.

Perhitungan akurasi klasifikasi dengan *confusion matrix* dapat digunakan untuk komposisi data yang *balance*. Akurasi klasifikasi pada data *unbalance* dapat dihitung menggunakan *geometric mean* (*G-mean*) dan *Area Under ROC Curve* (*AUC*). *G-mean* untuk kasus mengelompokan dengan jumlah kategori lebih dari dua dapat menggunakan *Expanding G-mean* dengan rumus sebagai berikut (Bekkar, Djemaa, & Alitouch, 2013).

$$G\text{-mean} = \left(\prod_{i=1}^g R_i \right)^{1/g} \quad (2.33)$$

$$AUC = \frac{1}{g} \sum_{i=1}^g R_i \quad (2.34)$$

Dimana

$$R_i = \frac{n_{ii}}{\sum_{l=1}^g n_{il}} \quad i = 1, 2, \dots, g$$

Keterangan:

n_{ii} : jumlah prediksi $\hat{\pi}_i$ yang tepat diklasifikasikan ke π_i

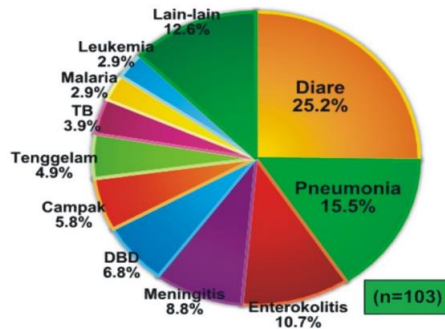
n_{il} : jumlah prediksi $\hat{\pi}_l$ yang salah diklasifikasikan ke π_i

2.6 Pneumonia

Pneumonia merupakan infeksi akut yang menyerang paru-paru atau alveoli yang disebabkan oleh mikroorganisme (PDPI, 2003). Definisi lain dari pneumonia adalah peradangan pada parenkim paru yang biasanya menyerang anak-anak atau balita dan secara klinis pneumonia terjadi sebagai penyakit primer atau komplikasi dari penyakit lain (Hockenberry & Wilson, 2009). UNICEF/WHO (2006) menyatakan pneumonia merupakan penyakit yang berasal dari infeksi akut di daerah saluran pernafasan bagian bawah yang secara spesifik mempengaruhi paru-paru. Ketika seseorang terkena pneumonia, alveoli terisi penuh oleh lendir dan cairan sehingga menyebabkan rasa sakit ketika bernafas dan oksigen yang terhirup akan menjadi sedikit.

Gambar 2.1 menunjukkan bahwa pneumonia merupakan penyakit penyebab kematian kedua tertinggi setelah diare pada balita. Hal ini menunjukkan bahwa pneumonia merupakan penyakit yang menjadi masalah pada kesehatan masyarakat utama yang berkontribusi terhadap tingginya angka kematian balita di Indonesia (Kemenkes RI, 2010).

Pneumonia dapat disebabkan oleh berbagai macam mikroorganisme, yaitu bakteri, virus, jamur, dan protozoa (PDPI, 2003). Diagnosis etiologi pneumonia pada balita sukar ditegakkan karena dahak sukar diperoleh. Sedangkan prosedur pemeriksaan imunologi belum memberikan hasil yang memuaskan untuk menentukan adanya bakteri penyebab pneumonia. Hanya biakan dari spesimen aspirasi paru serta pemeriksaan spesimen darah yang dapat diandalkan untuk membantu diagnosis etiologi pneumonia. Meskipun pemeriksaan spesimen fungsi paru merupakan cara yang sensitif untuk menentukan bakteri penyebab pneumonia pada balita, akan tetapi prosedur ini cukup berbahaya dan bertentangan dengan etika, terutama jika hanya dilakukan untuk penelitian



Sumber : Riskesdas 2007

Gambar 2.1 Proporsi Penyebab Kematian pada Umur 1-4 Tahun

Pneumonia dikelompokkan berdasarkan beberapa aspek, yaitu berdasarkan klinis dan epideologis, berdasarkan bakteri penyebab, dan berdasarkan predileksi infeksi (Rasyid & Kamso, 2004). Berdasarkan klinis dan epideologis, klasifikasi pneumonia dikelompokkan menjadi empat kelompok yaitu sebagai berikut.

- Pneumonia komunitas (*community-acquired pneumonia*), yaitu infeksi pneumonia yang terjadi di masyarakat.
- Pneumonia nosokomial (*hospital-acquires pneumonia / nosocomial pneumonia*), yaitu pneumonia apabila infeksi didapat di rumah sakit.
- Pneumonia aspirasi, adalah pneumonia yang disebabkan oleh terhirupnya bahan-bahan ke dalam saluran pernafasan.
- Pneumonia pada penderita *immunocompromised*.

Klasifikasi pneumonia berdasarkan bakteri penyebab dibagi menjadi empat kelompok yaitu sebagai berikut.

- Pneumonia bakterial/ tipikal. Pneumonia ini dapat terjadi pada semua usia. Beberapa bakteri mempunyai kecenderungan menyerang seseorang yang peka, misalnya bakteri *Klabsiella* pada penderita alkoholik atau bakteri *Staphylococcus* pada penderita pasca influenza.
- Pneumonia atipikal, disebabkan oleh *Mycoplasma*, *Legionella*, dan *Chlamydia*.
- Pneumonia virus.

- d. Pneumonia jamur yang biasanya merupakan infeksi sekunder. Pneumonia jenis ini terutama terjadi pada penderita dengan daya tahan lemah (*immunocompromised*).

Klasifikasi pneumonia berdasarkan prediksi infeksi dibagi menjadi tiga kelompok yaitu sebagai berikut.

- a. Pneumonia lobaris. Pneumonia yang terjadi pada satu lobus atau segmen kemungkinan disebabkan oleh obstruksi bronkus misalnya pada aspirasi benda asing.
- b. Bronkopneumonia. Pneumonia ini ditandai dengan munculnya bercak-bercak infiltrat pada lapangan paru. Dapat disebabkan baik oleh bakteri maupun virus. Bronkopneumonia sering terjadi pada bayi dan orang tua.

2.7 Diagnosis Pneumonia

Untuk melakukan diagnosis pada pasien pneumonia dilakukan dengan mengamati gejala klinis dan kelainan fisik melalui pemeriksaan fisik yang dilakukan oleh dokter. Pemeriksaan penunjang seperti *rontgen* dan pemeriksaan laboratorium masih diperlukan untuk memperkuat diagnosis apakah seseorang menderita pneumonia atau tidak. (Misnadiarly, 2008).

Sistem skor biasa digunakan untuk menentukan level risiko pasien penyakit pneumonia. Level risiko tersebut akan menjadi acuan dalam pengobatan dan perawatan terhadap pasien. Beberapa sistem skor yang biasa digunakan adalah sistem PSI (*Pneumonia Severity Index*) yang dikembangkan oleh PORT (*Pneumonia Patient Outcome Research Team*) dan sistem CURB-65 yang merupakan rekomendasi dari BTS (*British Thoracic Society*).

Sistem CURB-65 merupakan model skor yang sangat praktis, mudah diingat dan dinilai. Skor ini sudah divalidasi walaupun menggunakan sampel yang lebih sedikit dibanding PSI. Kelebihan skor ini adalah penggunaannya yang mudah dan dirancang untuk lebih menilai keparahan penyakit dibanding menilai pasien pneumonia dengan risiko mortalitas. Variabel-variabel yang terdapat dalam penentuan skor prediksi menggunakan sistem

CURB-65 dan jumlah poin tiap variabel dapat dilihat pada Tabel 2.2 sebagai berikut.

Tabel 2.2 Skor Prediksi CURB-65

Karakteristik	Poin
Penurunan Kesadaran	1
Urea nitrogren darah > 20 mg/dL	1
Frekuensi Pernapasan ≥ 30 per menit	1
Tekanan darah (sistolik < 90 mmHg atau diastolik ≤ 60 mmHg)	1
Usia ≥ 65 tahun	1

Kemudian skor CURB-65 tersebut dijumlahkan dan dikategorikan berdasarkan tingkat keparahannya seperti pada Tabel 2.3.

Tabel 2.3 Kelas Risiko dengan Sistem CURB-65

Jumlah Poin	Kelas Risiko	Rekomendasi Penanganan
0	Rendah	Rawat Jalan
1	Rendah	Rawat Jalan
2	Sedang	Rawat Inap / Rawat Jalan
3	Sedang ke Berat	Rawat Inap / Rawat Jalan
4 atau 5	Berat	Rawat Inap/ ICU

Persatuan Dokter Paru Indonesia (PDPI) menetapkan sistem PSI menjadi acuan diagnosis pneumonia. Menurut hasil *Pneumonia Patient Outcome Research Team* (PORT), beberapa karakteristik pasien dapat dimasukkan ke dalam kelas risiko seperti Tabel 2.4 sebagai berikut (PDPI, 2003).

Tabel 2.4 Sistem Skor pada Pneumonia Komunitas Sistem PSI

Karakteristik Penderita	Poin
Faktor demografi	
Usia: Laki-laki	Umur (tahun)
Perempuan	Umur (tahun)-10
Perawatan di rumah	+10
Penyakit penyerta	
Keganasan	+30
Penyakit hati	+20
Gagal jantung kongestif	+10
Penyakit serebrovaskuler	+10
Penyakit ginjal	+10

Tabel 2.4 (lanjutan)

Karakteristik Penderita	Poin
Pemeriksaan fisis	
Perubahan status mental	+20
Pernafasan > 30 kali/menit	+20
Takanan darah sistolik ≤ 90 mmHg	+20
Suhu tubuh < 35°C atau > 40°C	+15
Nadi ≥ 125 kali/menit	+10
Hasil laboratorium atau radiologi	
Analisis gas darah arteri : pH < 7,35	+30
BUN (<i>Blood urea nitrogen</i>) > 30 mg/dL	+20
Natrium < 130 mEq/liter	+20
Glukosa > 250 mg/dL	+10
Hematokrit < 30%	+10
PO ₂ ≤ 60 mmHg	+10
Efusi pleura	+10

Selanjutnya poin-poin dari hasil PORT dijumlahkan. Hasil penjumlahan tersebut kemudia dikategorikan menurut kelas risikonya, sehingga dapat ditentukan penanganan dan pengobatan yang tepat (PDPI, 2003).

Tabel 2.5 Derajat Skor Risiko dengan Sistem PSI

Risiko	Kelas	Jumlah Poin	Rekomendasi
Rendah	I	0	Rawat Jalan
	II	< 70	Rawat Jalan
	III	71 – 90	Rawat Inap / Rawat Jalan
Sedang	IV	91 – 130	Rawat Inap
Berat	V	> 130	Rawat Inap

(Halaman ini sengaja dikosongkan)

BAB III

METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Rumah Sakit Umum Daerah Dr. Soetomo. Data yang akan dianalisis adalah data rekam medis pasien pneumonia tahun 2015.

3.2 Kerangka Konsep

Penilaian kelas risiko digunakan untuk membantu diagnosis pneumonia. Penentuan kelas risiko dapat menggunakan sistem skoring CURB-65 maupun sistem PSI. Gambar 3.1 menunjukkan atribut-atribut yang merupakan kriteria penentuan kelas risiko baik menggunakan sistem CURB-65 maupun sistem PSI (PDPI, 2003).

Penelitian ini dilakukan untuk mengkaji metode analisis diskriminan dan naïve bayes dalam mengklasifikasikan kelas risiko pneumonia, sehingga atribut penentuan kelas risiko pneumonia yang digunakan sebagai variabel independen adalah atribut dengan skala data rasio.

3.3 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah sebagai berikut.

Tabel 3.1 Variabel Penelitian

Variabel	Keterangan	Skala
Y	Kelas Risiko Pneumonia: II/III/IV/V	Ordinal
X ₁	Usia (tahun)	Rasio
X ₂	Tekanan Darah Sistolik (mmHg)	Rasio
X ₃	Tekanan Darah Diastolik (mmHg)	Rasio
X ₄	Denyut Nadi (denyut/menit)	Rasio
X ₅	Frekuensi Nafas tiap menit	Rasio
X ₆	Suhu tubuh (°C)	Rasio
X ₇	Analisis Gas Darah Arteri (pH)	Rasio
X ₈	Kadar Natrium dalam darah (mEq/liter)	Rasio
X ₉	Kadar Glukosa dalam darah(mg/dL)	Rasio

Tabel 3.1 (lanjutan)

Variabel	Keterangan	Skala
X ₁₀	Persentase sel darah merah terhadap volume darah /Hematokrit (%)	Rasio
X ₁₁	Tekanan oksigen dalam darah /Kadar PO ₂ (mmHg)	Rasio
X ₁₂	Urea Nitrogen Darah (mg/dL)	Rasio

3.4 Struktur Data Penelitian

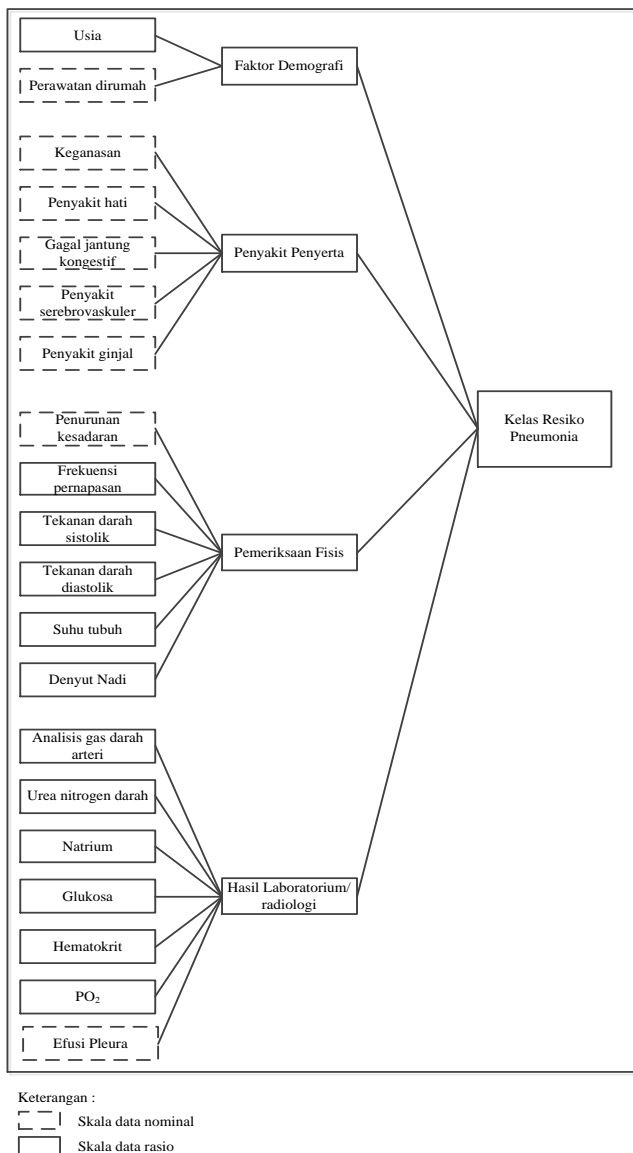
Struktur data yang digunakan dalam penelitian ini disajikan pada Tabel 3.2.

Tabel 3.2 Struktur Data

Sample	Respon	Variabel Prediktor			
	Y	X ₁	X ₂	...	X _p
1	y ₂₁	x ₁₁	x ₁₂	...	x _{1p}
2	y ₂₂	x ₂₁	x ₂₂	...	x _{2p}
⋮	⋮	⋮	⋮	⋮	⋮
n ₂	y _{2n₂}	x _{n₂1}	x _{n₂2}	...	x _{n₂p}
1	y ₃₁	x ₁₁	x ₁₂	...	x _{1p}
2	y ₃₂	x ₂₁	x ₂₂	...	x _{2p}
⋮	⋮	⋮	⋮	⋮	⋮
n ₃	y _{3n₃}	x _{n₃1}	x _{n₃2}	...	x _{n₃p}
1	y ₄₁	x ₁₁	x ₁₂	...	x _{1p}
2	y ₄₂	x ₂₁	x ₂₂	...	x _{2p}
⋮	⋮	⋮	⋮	⋮	⋮
n ₄	y _{4n₄}	x _{n₄1}	x _{n₄2}	...	x _{n₄p}
1	y ₅₁	x ₁₁	x ₁₂	...	x _{1p}
2	y ₅₂	x ₂₁	x ₂₂	...	x _{2p}
⋮	⋮	⋮	⋮	⋮	⋮
n ₅	y _{5n₅}	x _{n₅1}	x _{n₅2}	...	x _{n₅p}

Keterangan:

- n₂ : Jumlah observasi pada kategori kelas resiko II
- n₃ : Jumlah observasi pada kategori kelas resiko III
- n₃ : Jumlah observasi pada kategori kelas resiko IV
- n₅ : Jumlah observasi pada kategori kelas resiko V



Gambar 3.1 Kriteria Penentuan Kelas Risiko Pneumonia

3.5 Langkah Analisis

Langkah analisis yang dilakukan dalam penelitian ini adalah sebagai berikut.

1. Melakukan analisis statistika deskriptif dan eksplorasi data.
2. Membagi data menjadi data *training* dan *testing* dengan proporsi 80:20. Proporsi variabel independen untuk setiap kelas adalah sama.
3. Melakukan klasifikasi menggunakan analisis diskriminan linier, dengan metode seleksi variabel yang digunakan adalah metode *stepwise*, *forward selection*, *backward elimination*, dan *Particle Swarm Optimization* (PSO) pada data *training*.
 - i. *Forward selection* dilakukan dengan memasukkan variabel independen satu per satu berdasarkan kriteria nilai *Wilks' A*.
 - ii. *Backward elimination* dilakukan dengan mengeliminasi variabel independen satu per satu berdasarkan kriteria nilai *Wilks' A*.
 - iii. *Stepwise method* dilakukan dengan memasukkan variabel independen menggunakan metode *forward selection*, apabila variabel independen yang dimasukan menaikan nilai *Wilks' A* maka variabel tersebut dieliminasi menggunakan *backward elimination*.
 - iv. Langkah-lahkah seleksi variabel dalam *Particle Swarm Optimization* (PSO) adalah sebagai berikut.
 - 1) Menentukan posisi dan kecepatan setiap partikel secara random.
 - 2) Mengumpulkan variabel yang terpilih oleh partikel, kemudian menghitung nilai *Fitness* dari variabel yang terpilih dengan menggunakan akurasi dari klasifikasi
 - 3) Jika *fitness* partikel lebih besar dari *fitness* partikel terbaik (*pbest*) maka *update* nilai *pbest*. Jika *fitness*

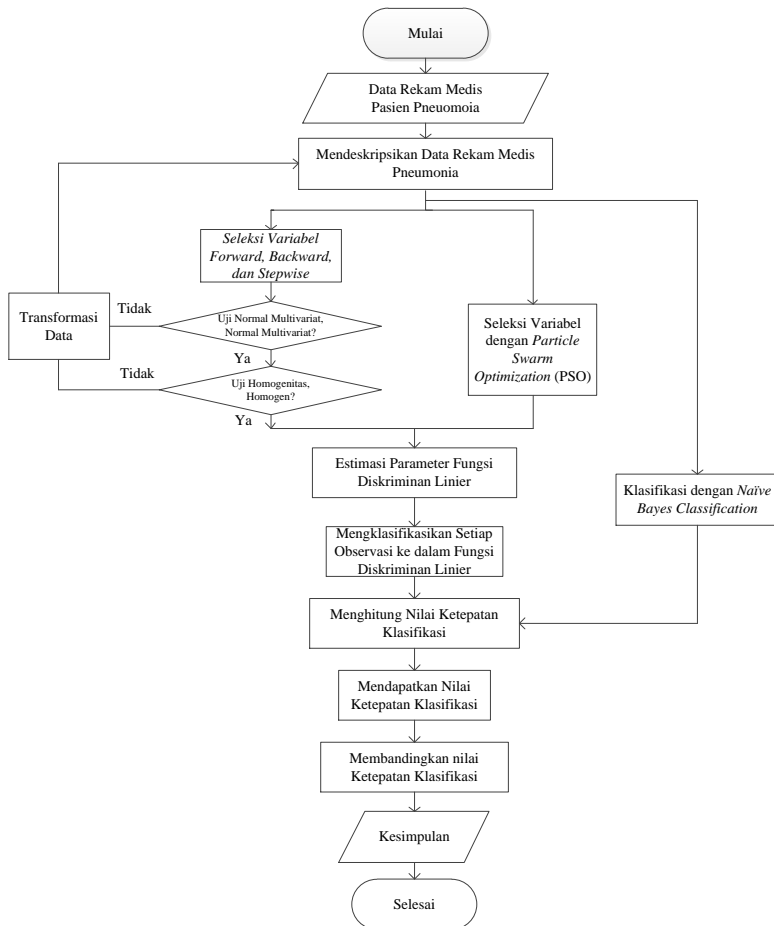
- partikel lebih besar dari pada *fitness* terbaik dari keseluruhan (*gbest*) maka *update* nilai *gbest*.
- 4) *Update* posisi dan kecepatan setiap partikel berdasarkan nilai *pbest* dan *gbest* terakhir.
 - 5) Apabila nilai *fitness* sudah konvergen maka variabel yang dipilih sudah optimum.
- a. Melakukan deteksi multikolinieritas antar variabel independen.
 - b. Melakukan uji distribusi normal multivariat untuk seluruh variabel independen.
 - c. Melakukan uji homogenitas matriks varian-kovarian.
 - d. Mengestimasi koefisien parameter fungsi diskriminan.
 - e. Menghitung ketepatan klasifikasi.
 - f. Melakukan validasi akhir.
4. Melakukan klasifikasi dengan menggunakan *Naïve Bayes Classification* baik dengan mengikutsertakan semua variabel independen maupun variabel independen yang telah terseleksi menggunakan *forward selection*, *backward elimination*, *stepwise method*, dan *Particle Swarm Optimizatin*. Langkah-langkah *Naïve Bayes Classification* adalah sebagai berikut.
- a. Menghitung nilai probabilitas dari masing-masing parameter pada setiap kategori. Apabila data adalah numerik (kontinyu), maka menentukan nilai rata-rata, varian dan standar deviasi dari masing-masing parameter setiap kategori kemudian menentukan nilai probabilitas menggunakan pendekatan distribusi normal.
 - b. Menentukan probabilitas akhir dari semua parameter terhadap setiap kategori.
 - c. Menentukan kelompok kategori berdasarkan nilai probabilitas yang tertinggi.
 - d. Menghitung ketepatan hasil klasifikasi.
5. Membandingkan variabel yang terpilih dari hasil seleksi variabel dan nilai ketepatan klasifikasi pada data *testing* hasil analisis metode Analisis Diskriminan Linier, *hybrid*

Analisis Diskriminan Linier – *Particle Swarm Optimization* (ADL-PSO) dan *Naïve Bayes Classification*.

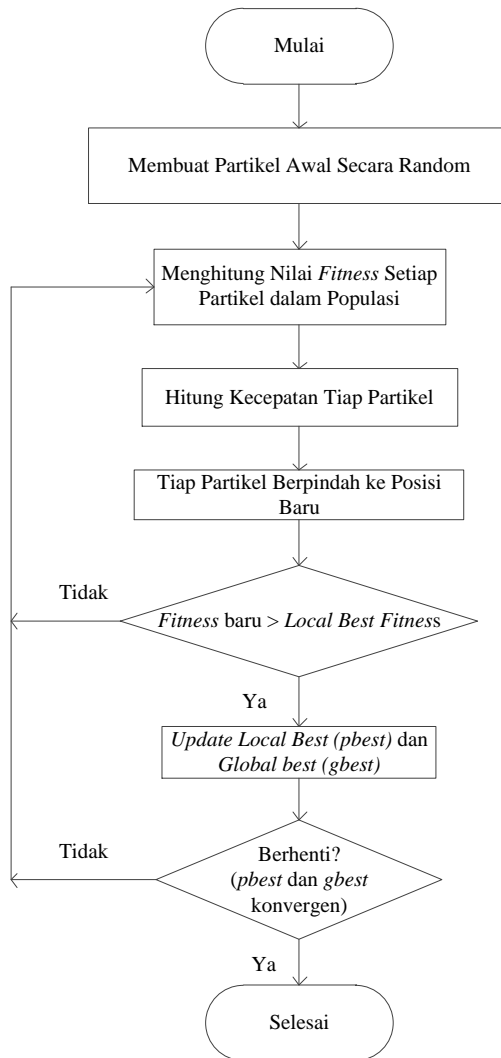
6. Menarik kesimpulan dan memberi saran.

3.6 Diagram Alir

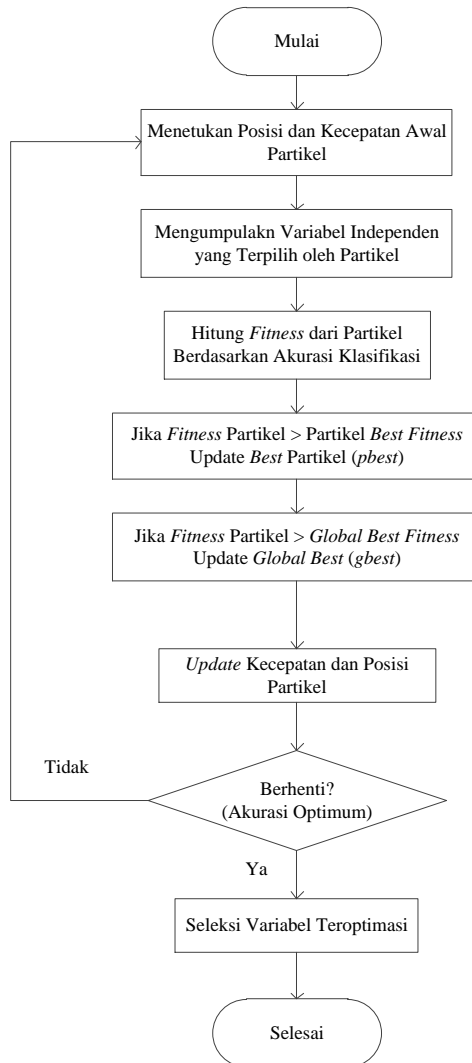
Diagram alir menggambarkan proses analisis data dari awal hingga selesai. Diagram alir dalam penelitian ini ditunjukkan pada Gambar 3.2, Gambar 3.3, Gambar 3.4, dan Gambar 3.5.



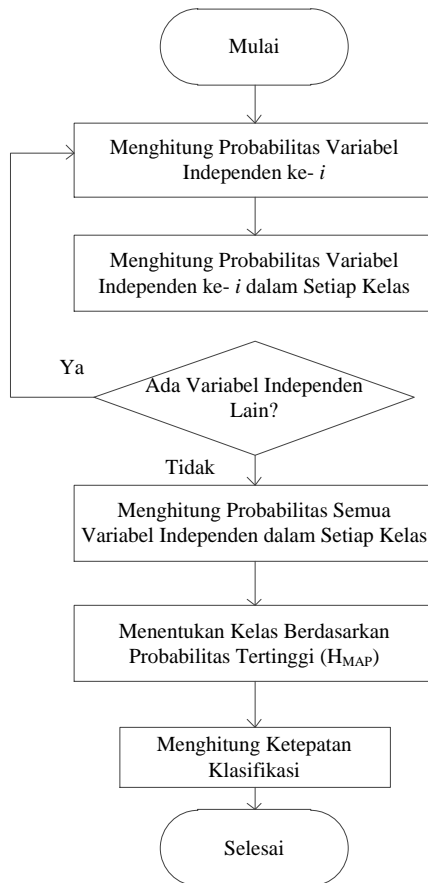
Gambar 3.2 Diagram Alir Penelitian



Gambar 3.3 Diagram Alir PSO



Gambar 3.4 Diagram Alir Seleksi Variabel PSO



Gambar 3.5 Diagram Alir Klasifikasi Naïve Bayes

BAB IV

ANALISIS DAN PEMBAHASAN

Pada bab ini akan dibahas mengenai hasil klasifikasi menggunakan analisis diskriminan linier, analisis diskriminan linier dengan *particle swarm optimization* dan klasifikasi naive bayes.

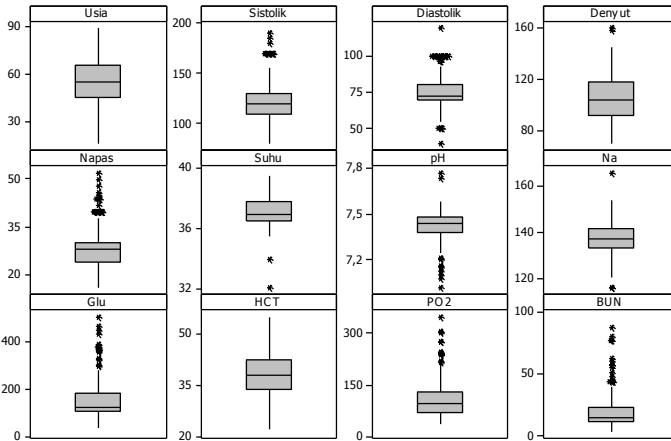
4.1 Karakteristik Data Pasien Pneumonia di RSUD Dr. Soetomo

Sebelum melakukan klasifikasi perlu diketahui perbedaan karakteristik setiap variabel yang diduga berpengaruh terhadap penentuan kelas risiko penyakit pneumonia. Data rekam medis pasien pneumonia terdiri atas 12 variabel independen, 4 kategori pada variabel dependen, dan terdapat 186 observasi. Karakteristik data yang dianalisis yaitu rata-rata, standar deviasi, nilai minimum, dan nilai maksimum. Karakteristik data rekam medis pasien pneumonia disajikan dalam Tabel 4.1.

Tabel 4.1 Statistika Deskriptif Variabel Independen

Variabel	Mean	StDev	Min	Max
Usia (tahun)	55,28	14,26	16,00	89,00
Sistolik (mmHg)	119,88	20,82	80,00	190,00
Diastolik (mmHg)	74,49	12,05	40,00	119,00
Nadi (kali/menit)	105,17	16,45	70,00	160,00
Napas (kali/menit)	28,02	6,51	16,00	52,00
Suhu (°C)	37,08	0,93	32,10	39,50
pH	7,42	0,10	7,02	7,77
Natrium (mEq/L)	136,68	7,02	115,90	166,00
Glukosa (mg/dL)	156,78	84,20	32,00	503,00
Hematokrit (%)	38,03	6,49	21,80	54,90
PO ₂ (mmHg)	107,84	55,57	33,400	349,00
BUN (mg/dL)	19,59	14,56	3,00	88,00

Semua variabel yang diduga mempengaruhi penentuan kelas risiko merupakan hasil pemeriksaan pasien pneumonia saat pertama kali masuk di RSUD Dr. Soetomo. Untuk mengetahui lebih jelas karakteristik variabel tersebut, disajikan pula dalam bentuk *boxplot* dalam Gambar 4.1 sebagai berikut.



Gambar 4.1 Karakteristik Variabel Independen

Tabel 4.1 menunjukkan rata-rata penderita pneumonia berada di usia lanjut yang ditunjukkan dengan nilai rata-rata usia sebesar 55,28 tahun, dengan pasien termuda yaitu usia 16 tahun dan pasien yang paling tua berusia 89 tahun. Hanya terdapat sedikit yang memiliki usia terlalu muda atau terlalu tua atau dapat dikatakan variabel usia memiliki distribusi yang simetris, hal dapat dilihat dari *boxplot* variabel usia pada Gambar 4.1 yang simetris. Rata-rata tekanan darah sistolik pasien pneumonia masih dikatakan normal karena bernilai lebih dari 90 mmHg, yaitu sebesar 119,88 mmHg, akan tetapi masih ada pasien yang memiliki tekanan sistolik yang sangat rendah dan rawan dikatakan berbahaya bagi pasien pneumonia yaitu

hanya sebesar 80 mmHg dan tekanan sistolik pasien pneumonia tetinggi yaitu sebesar 190 mmHg dimana tekanan sistolik tersebut dikatakan sangat tinggi akan tetapi tidak dianggap berbahaya bagi penderita pneumonia. Gambar 4.1 menunjukkan bahwa *boxplot* tekanan darah sistolik berbentuk simetris, hal ini berarti bahwa data pasien pneumonia memiliki distribusi yang simetris atau dengan kata lain hanya terdapat sedikit pasien yang memiliki tekanan darah sistolik terlalu tinggi atau terlalu rendah.

Rata-rata tekanan darah diastolik pasien pneumonia sudah lebih dari 60 mmHg sehingga masih dikatakan normal yaitu sebesar 74,495 mmHg, tekanan diastolik pasien pneumonia terendah berada di bawah normal yaitu 40 mmHg dan tekanan diastolik tertinggi adalah 119 mmHg, sama seperti tekanan sistolik, walaupun dikatakan sangat tinggi akan tetapi tidak berbahaya bagi penderita pneumonia. Berbeda dengan tekanan darah sistolik, tekanan darah diastolik memiliki distribusi *skewness* kanan. Hal ini dapat dilihat dari *boxplot* tekanan darah diastolik yang tidak simetris. Nilai rata-rata lebih besar dari median, yang berarti kebanyakan data adalah data dengan nilai yang kecil.

Rata-rata denyut nadi pasien pneumonia yaitu sebesar 105,1, rata-rata tersebut masih dikatakan normal karena masih berada di bawah 125 kali tiap menit. Denyut nadi terendah pada pasien adalah 60 kali tiap menit dan denyut nadi tertinggi adalah 190 kali tiap menit, dimana angka tersebut dikatakan sangat tinggi dan berbahaya bagi pasien pneumonia. Data denyut nadi pasien pneumonia memiliki distribusi yang simetris, hal ini ditunjukkan dari bentuk *boxplot* yang simetris. Hanya terdapat beberapa data yang bernilai sangat besar atau sangat kecil, atau kebanyakan data berada di sekitar nilai median.

Rata-rata frekuensi napas pasien pneumonia masih dikatakan normal yaitu sebesar 28,022 kali tiap menit. Hanya

sekitar kurang dari 20% yang bernilai lebih dari median, hal ini menunjukkan data memiliki distribusi *skewness* kiri dan terdapat beberapa data yang berada diluar batas atas *boxplot* dan lebih dari 30 kali napas tiap menit. Frekuensi napas terendah pada pasien pneumonia adalah 16 kali dan frekuensi napas tertinggi adalah 52 kali tiap menit, tinggi nya frekuensi napas merupakan salah satu indikasi bahwa pneumonia pasien tersebut sudah cukup parah. Suhu tubuh pasien memiliki rata-rata yang normal yaitu 37,076°C. Dari *boxplot* pada variabel suhu menunjukkan bahwa hampir semua pasien memiliki suhu tubuh yang masih normal yaitu bernilai antara 35°C - 40°C. Distribusi data yang ditunjukkan oleh *boxplot* adalah *skewness* kanan, yang berarti kebanyakan pasien memiliki suhu tubuh tinggi. Suhu tubuh terendah berada di bawah batas normal yaitu hanya sebesar 32,1°C, akan tetapi suhu tubuh tertinggi pasien masih dikatakan normal yaitu sebesar 39,5°C.

Analisis gas darah arteri pasien memiliki rata-rata yang masih normal, yaitu sebesar 7,42. Kebanyakan data analisis gas darah pasien memiliki pH yang rendah, hal ini ditunjukkan dari *boxplot* yang membentuk distribusi *skewness* kiri. Sedangkan analisis gas darah pasien terendah sebesar 7,02 dimana pH tersebut masih dibawah normal karena kurang dari 7,35 dan analisis gas darah tertinggi sebesar 7,77. Kadar natrium darah pasien pneumonian memiliki rata-rata yang masih dikatakan normal yaitu sebesar 136,68 mEq/liter. Hanya terdapat sedikit pasien yang memiliki kadar natrium darah yang sangat tinggi atau sangat rendah, hal ini ditunjukkan dari bentuk *boxplot* yang simetris. Kadar natrium pasien tertinggi sebesar 166 mEq/liter dan kadar natrium terendah sebesar 115,6, dimana kadar natrium tersebut dikatakan sangat rendah dan berada di bawah batas normal 130 mEq/liter.

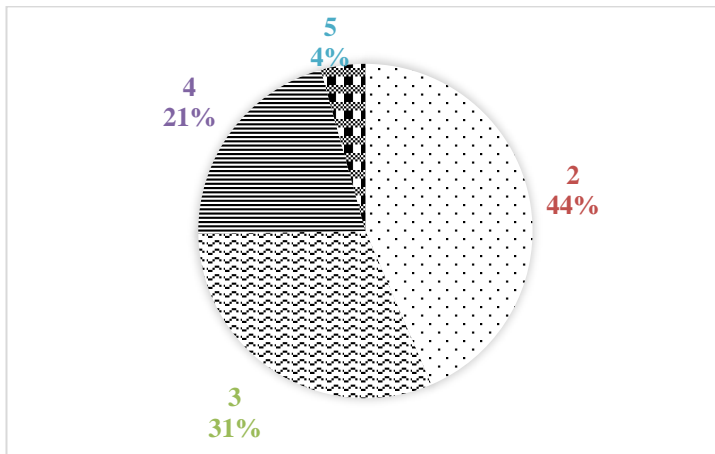
Rata-rata kadar gula dalam darah pasien pneumonia masih dikatakan normal yaitu sebesar 156,78 mg/liter. Kebanyakan

pasien memiliki kadar glukosa yang cukup tinggi, hal ini ditunjukkan dari *boxplot* yang berbentuk *skweness* kanan. Kadar glukosa darah terendah yaitu sebesar 32 mg/liter, angka ini cukup kecil akan tetapi pada kasus penyakit pneumonia tidak terlalu menjadi perhatian. Sedangkan kadar gula darah pasien yang tertinggi yaitu sebesar 503 mg/liter, angka tersebut dikatakan sangat tinggi karena lebih dari dua kali lipat batas normal 250 mg/liter dan hal ini merupakan salah satu hal yang menjadi indikasi tingkat keparahan penyakit pneumonia. Persentase sel darah merah dibandingkan dengan volume darah atau biasa disebut hematokrit pada pasien pneumonia memiliki rata-rata sebesar 38,03%. Persentase hematokrit tertinggi sebesar 54,9% dan persentase hematokrit terendah adalah 21,8%, dimana persentase tersebut dipandang cukup rendah dan dianggap berbahaya bagi pasien pneumonia karena bernilai kurang dari 30%. Hanya terdapat sedikit pasien yang memiliki persentase kematokrit yang sangat rendah atau sangat tinggi, hal ini dapat dilihat dari *boxplot* yang simetris.

Tekanan oksigen dalam darah atau kadar PO_2 pasien pneumonia memiliki rata-rata sebesar 107,84 mmHg, dimana kadar PO_2 tersebut masih dikatakan normal. Kadar PO_2 pasien terendah adlah 33,4 mmHg, dimana kadar tersebut berada di bawah normal dan menjadi salah satu pertimbangan penentuan keparahan pneumonia. Sedangkan kadar PO_2 tertinggi sebesar 349 mmHg, walaupun angka tersebut cukup tinggi tapi tidak menjadi pertimbangan dalam penentuan keparahan pneumonia. kebanyakan pasien memiliki kadar PO_2 yang tinggi, dilihat dari *boxplot* yang berbentuk *skewness* kanan. Rata-rata BUN (*Blood Urea Nitrogen*) pada pasien pneumonia adalah sebesar 19,59 mg/dL, dimana rata-rata tersebut masih dikatakan normal karena tidak labih dari 30 mg/dL. Kebanyakan pasien memiliki BUN yang tinggi, dilihat dari *boxplot* yang membentuk pola distribusi *skewness* kanan. BUN pasien terendah yaitu 3 mg/dL

dan BUN pasien tertinggi yaitu sebesar 88 mm/dL, angka tersebut sangat jauh diatas normal karena bernilai hampir tiga kali lipat dari batas yang seharusnya dan dianggap cukup berbahaya bagi pasien pneumonia.

Kemudian disajikan diagram mengenai proporsi tiap kategori pada variabel kelas risiko pasien pneumonia yang dapat dilihat pada Gambar 4.1. Kelas risiko pasien ditentukan menggunakan sistem PSI.



Gambar 4.2 Proporsi Tiap Kelas Risiko Pneumonia

Gambar 4.2 menunjukkan tiap kelas risiko pasien pneumonia memiliki proporsi yang berbeda (*unbalance*) dan hanya terdapat empat kategori kelas risiko yaitu kelas II, III, IV dan V. Tidak ada pasien yang masuk dalam kategori kelas risiko tipe I. Kelas risiko tipe II memiliki proporsi sebesar 44% atau sejumlah 81 dari total 186 pasien. Untuk kelas risiko tipe III memiliki proporsi sebesar 31% atau sejumlah 58 dari keseluruhan pasien. Proporsi kelas risiko tipe IV sebesar 21% atau sejumlah 39 dari keseluruhan pasien pneumonia. Sedangkan kelas risiko tipe V memiliki proporsi yang sangat

kecil dibanding kelas lainnya, yaitu sebesar 4% atau sejumlah 8 pasien dari total 186 pasien pneumonia.

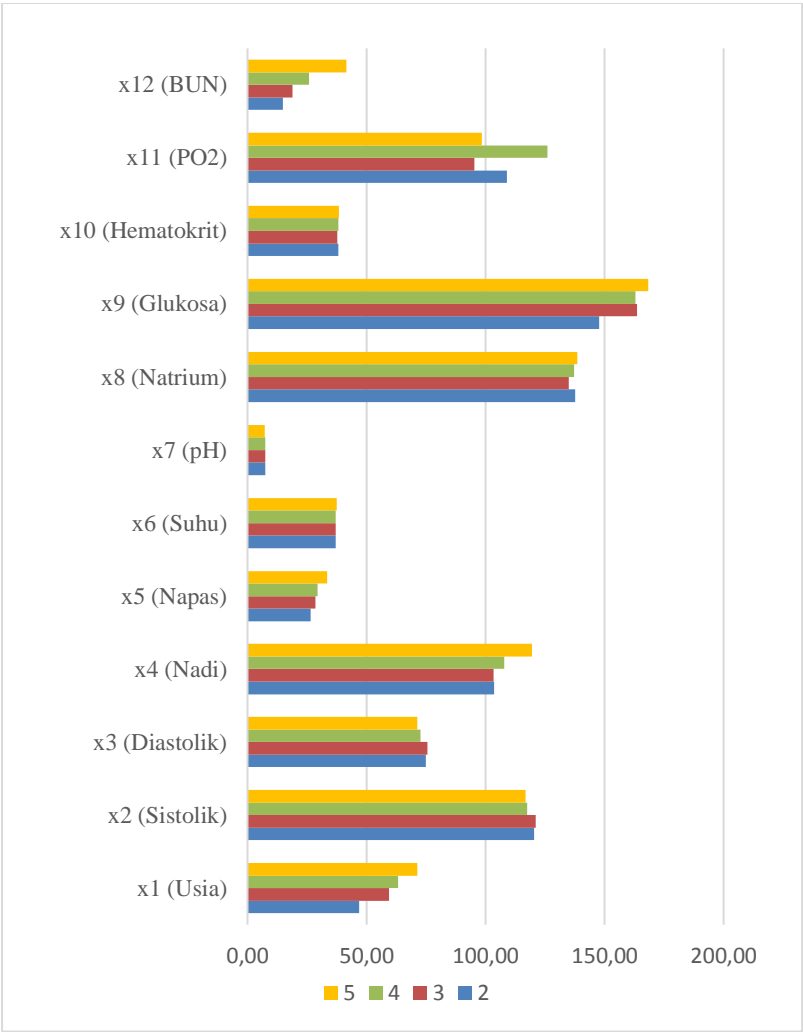
Statistika deksriptif variabel yang diduga berpengaruh terhadap penentuan kelas risiko penyakit pneumonia tiap kategori kelas risiko ditunjukkan dalam Tabel 4.2 sebagai berikut.

Tabel 4.2 Statistika Deskriptiv Variabel Independen Tiap Kategori

	<i>Mean</i>				<i>StDev</i>			
	K2	K3	K4	K5	K2	K3	K4	K5
x_1	46,88	59,40	63,33	71,38	12,37	12,48	10,73	10,90
x_2	120,43	121,16	117,44	116,88	19,49	18,07	26,25	25,49
x_3	74,99	75,52	72,62	71,25	10,39	12,33	14,47	13,56
x_4	103,69	103,47	107,85	119,50	15,09	17,67	16,25	15,82
x_5	26,42	28,55	29,44	33,50	5,32	6,98	7,30	5,73
x_6	37,04	37,04	37,11	37,58	0,98	0,82	0,97	1,01
x_7	7,45	7,43	7,36	7,34	0,08	0,09	0,12	0,09
x_8	137,54	134,93	137,11	138,65	5,40	7,04	8,98	9,47
x_9	147,74	163,70	162,90	168,40	79,34	90,10	86,00	86,40
x_{10}	38,21	37,69	38,06	38,50	5,80	6,81	7,07	8,76
x_{11}	109,08	95,24	126,00	98,38	56,72	39,05	72,40	26,84
x_{12}	14,88	18,97	25,80	41,50	10,69	13,03	16,06	22,05

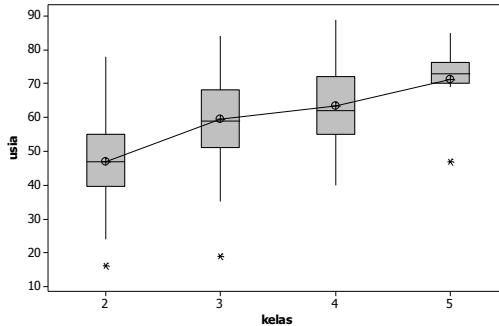
Gambar 4.3 menunjukkan diagram batang setiap variabel yang diduga berpengaruh terhadap penentuan kelas risiko pasien pneumonia berdasarkan kategori kelas risiko. Berdasarkan Tabel 4.2 dan Gambar 4.3 menunjukkan perbedaan rata-rata variabel yang cukup signifikan pada tiap kategori kelas risiko pada variabel x_1, x_4, x_9, x_{11} , dan x_{12} . Sedangkan variabel x_2, x_3, x_5 dan x_8 masih terdapat perbedaan

rata-rata namun tidak jauh berbeda. Pada variabel x_6, x_7 , dan x_{10} memiliki rata-rata yang hampir sama tiap kategori.



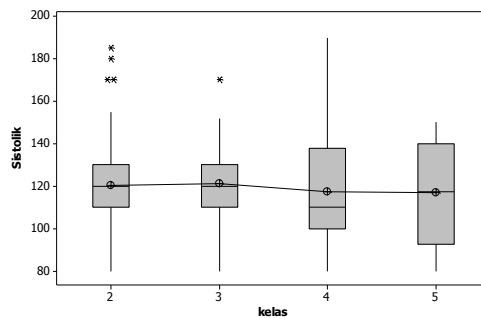
Gambar 4.3 Perbandingan Rata-Rata Variabel Tiap Kategori

Untuk melihat perbedaan rata-rata tiap variabel independen secara lebih jelas dapat dilihat pada Gambar 4.4 hingga Gambar 4.15 sebagai berikut.



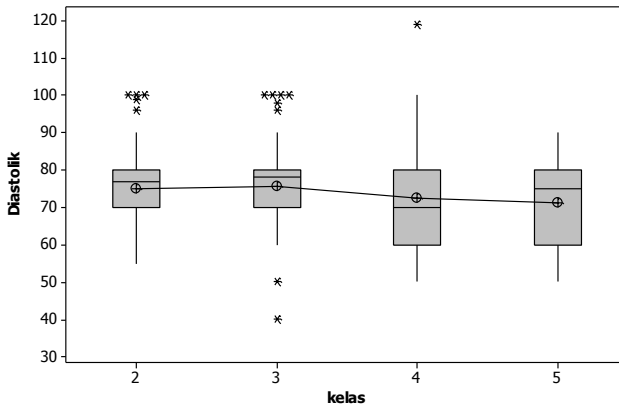
Gambar 4.4 Perbedaan Rata-Rata Variabel Usia Antar Kategori

Gambar 4.4 menunjukkan *boxplot* usia pasien pneumonia berdasarkan kelas risiko. Rata-rata usia berbeda tiap kelas risiko, ditunjukkan oleh garis *mean* pada *boxplot* yang tidak sejajar atau berada pada titik yang sama. Kategori kelas risiko 2,3 dan 5 menunjukkan distribusi yang simetris, sedangkan pada kelas risiko 4 lebih banyak pasien yang berusia tua dibanding usia muda yang ditunjukkan dari *boxplot* yang berbentuk *skewness kanan*



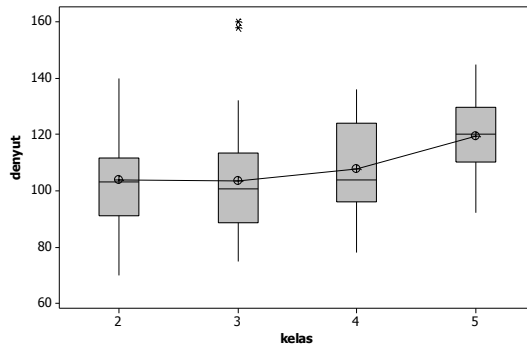
Gambar 4.5 Perbedaan Rata-Rata Tekanan Darah Sistolik Antar Kategori

Gambar 4.5 menunjukkan bahwa rata-rata tekanan darah sistolik pasien pneumonia tiap kelas risiko tidak berbeda secara signifikan. Hal tersebut dapat dilihat dari garis *mean* yang hampir sejajar. Hanya terjadi sedikit perbedaan rata-rata tiap kategori kelas risiko, terutama pada kategori kelas risiko 3 dan kelas risiko 4. Nilai rata-rata dan median tekanan darah sistolik pasien hampir sama kecuali kategori kelas risiko 4. Hal ini disebabkan karena pada kategori 4 pasien yang memiliki tekanan sistolik tinggi lebih banyak dibanding tekanan sistolik rendah.



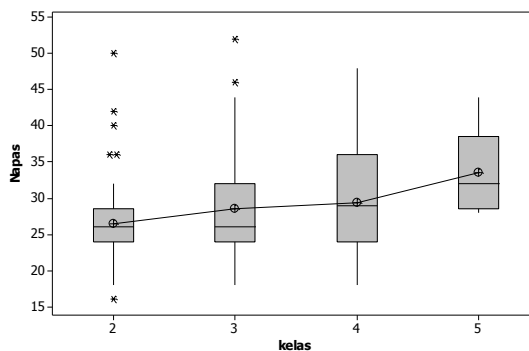
Gambar 4.6 Perbedaan Rata-Rata Tekanan Darah Diastolik

Gambar 4.6 menunjukkan bahwa terdapat perbedaan rata-rata tekanan darah diastolik pada tiap kategori kelas risiko. Hal ini ditunjukkan dari garis *mean* yang tidak menunjukkan garis sejajar. Nilai mean dan median tekanan darah diastolik tiap kategori kelas risiko tidak sama namun tidak jauh berbeda. Hal ini disebabkan terdapat data tekanan darah diastolik yang berbeda cukup jauh. .



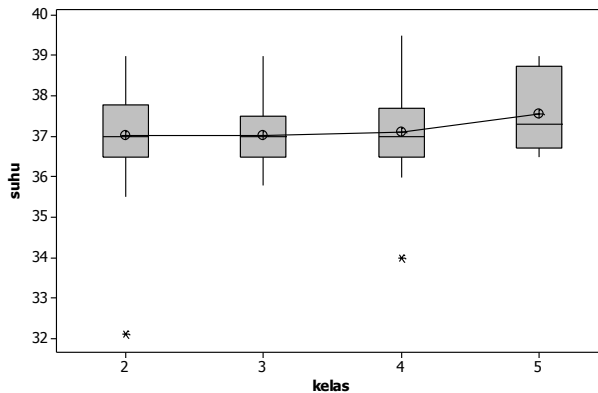
Gambar 4.7 Perbedaan Rata-Rata Denyut Nadi Antar Kategori

Gambar 4.7 menunjukkan *boxplot* denyut nadi pasien pneumonia tiap kategori kelas risiko. Dapat dilihat rata-rata denyut nadi tiap kategori kelas risiko berbeda, hal ini ditunjukkan dari garis mean yang tidak sejajar. Rata-rata dan median pada kategori kelas risiko 2 dan 5 sama namun pada kategori kelas risiko 3 dan 4. Perbedaan rata-rata dan median kelas risiko 3 dan 4 disebabkan karena terdapat angka yang sangat berbeda.



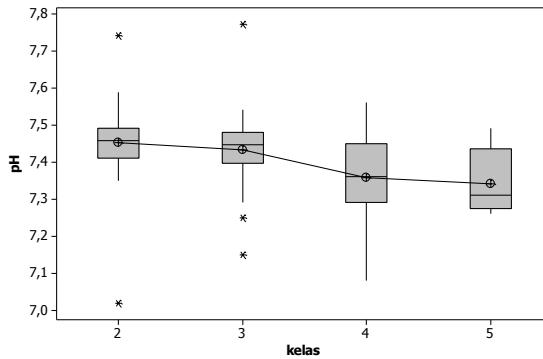
Gambar 4.8 Perbedaan Rata-Rata Frekuensi Napas Antar Kategori

Gambar 4.8 merupakan *boxplot* frekuensi napas pasien pneumonia tiap kategori kelas risiko. Dapat dilihat bahwa terdapat perbedaan rata-rata frekuensi napas tiap kategori kelas risiko pneumonia. Hal ini dapat dilihat dari garis *mean* yang tidak sejajar dan cenderung naik pada setiap kategori kelas risiko. Nilai median dan rata-rata frekuensi napas tiap kategori hampir sama kecuali pada kategori kelas risiko 3 dan kelas risiko 5, perbedaan nilai median dan rata-rata tersebut disebabkan karena terdapat data ekstrem atau bernilai jauh berbeda dari data-data lainnya.



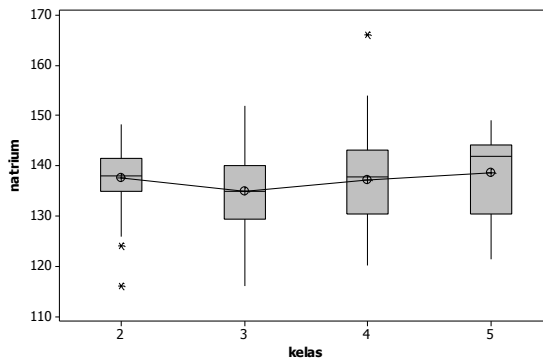
Gambar 4.9 Perbedaan Rata-Rata Suhu Tubuh Antar Kategori

Gambar 4.9 menunjukkan *boxplot* suhu tubuh pasien pneumonia tiap kategori kelas risiko. Terdapat perbedaan rata-rata suhu tubuh pasien tiap kategori walaupun tidak berbeda secara signifikan. Rata-rata suhu tubuh pada kategori kelas risiko 1 dan kelas risiko 2 hampir sama, kemudian berbeda mulai kelas risiko 3 hingga kelas risiko 4. Nilai rata-rata dan median pada tiap kelas risiko hampir sama.



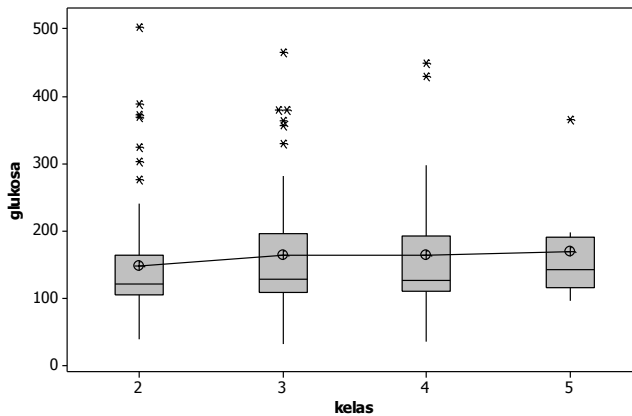
Gambar 4.10 Perbedaan Rata-Rata Analisis Gas Darah Arteri Antar Kategori

Gambar 4.10 merupakan *boxplot* variabel analisis gas darah arteri (dalam pH) pasien pneumonia tiap kategori kelas risiko. Dapat dilihat bahwa rata-rata analisis gas darah arteri pasien berbeda tiap kategori kelas risiko, yang ditunjukkan dengan garis *mean* yang tidak sejajar. Nilai median dan mean pada tiap kategori hampir sama.



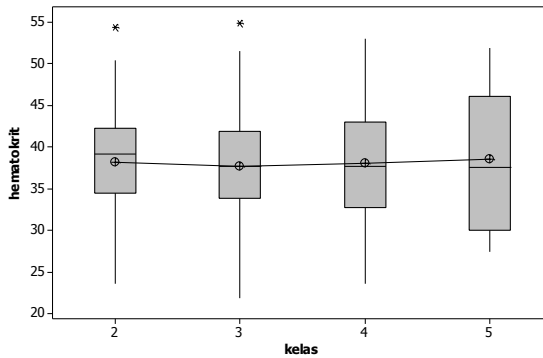
Gambar 4.11 Perbedaan Rata-Rata Kadar Natrium Darah Antar Kategori

Gambar 4.11 merupakan *boxplot* kadar natrium dalam darah pasien pneumonia tiap kategori kelas risiko. Rata-rata kadar natrium dalam darah tiap kelas risiko berbeda walaupun tidak secara signifikan. Nilai median dan rata-rata kadar natrium darah hampir sama pada setiap kategori kelas risiko kecuali pada kelas risiko 5. Hal ini disebabkan karena terdapat data yang ekstrim pada kategori tersebut.



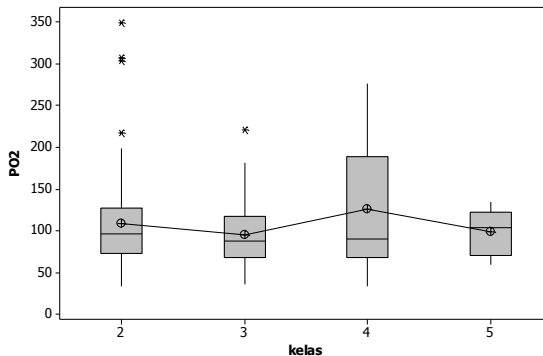
Gambar 4.12 Perbedaan Rata-Rata Kadar Glukosa Darah Antar Kategori

Boxplot kadar glukosa dalam darah pasien pneumonia pada tiap kategori kelas risiko ditunjukkan oleh Gambar 4.12. Rata-rata kadar glukosa dalam darah pada tiap kategori kelas risiko tidak berbeda secara signifikan, pada kelas risiko 3 dan kelas risiko 4 menunjukkan rata-rata yang sama. Hanya terjadi sedikit perubahan rata-rata pada kategori kelas risiko 2 dan kelas risiko 5. Nilai median dan rata-rata pada tiap kelas risiko berbeda, hal ini disebabkan adanya nilai ekstrem pada tiap kategori.



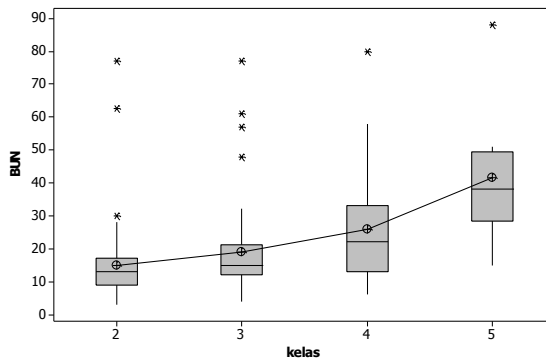
Gambar 4.13 Perbedaan Rata-Rata Hematokrit Antar Kategori

Gambar 4.13 menunjukkan *boxplot* persentase hematokrit atau perbandingan sel darah merah terhadap volume darah tiap kategori kelas risiko. Dapat dilihat bahwa tidak terdapat perbedaan rata-rata persentase hematokrit yang signifikan. Hal ini dapat dilihat dari garis *mean* yang hampir sejajar. Nilai rata-rata dan median dari persentase hematokrit tiap kategori kelas risiko hampir sama.



Gambar 4.14 Perbedaan Rata-Rata Kadar PO₂ Antar Kategori

Gambar 4.14 menunjukkan *boxplot* kadar PO_2 pasien pneumonia tiap kategori kelas risiko. Garis *mean* pada *boxplot* tidak sejajar atau tidak segaris, hal ini menunjukkan bahwa rata-rata kadar PO_2 tiap kategori kelas risiko berbeda. Nilai rata-rata dan median pada tiap kategori hampir sama kecuali pada kategori kelas risiko 2 dan kelas risiko 4. Hal ini disebabkan karena adanya data ekstrem pada kategori tersebut.



Gambar 4.15 Perbedaan Rata-Rata Variabel BUN Antar Kategori

Gambar 4.14 menunjukkan *boxplot* variabel BUN (*Blood Urea Nitrogen*) pasien pneumonia tiap kategori kelas risiko. Dapat dilihat bahwa rata-rata BUN tiap kategori kelas risiko berbeda secara signifikan, hal ini ditunjukkan dari garis *mean* yang tidak sejajar dan cenderung meningkat. Nilai rata-rata dan median berbeda tiap kategori kelas risiko. Sebagian besar variabel independen memiliki perbedaan rata-rata yang cukup signifikan sehingga secara visual dapat dilanjutkan dengan analisis diskriminan.

4.2 Analisis Diskriminan Pasien Pneumonia

Langkah pertama dalam analisis diskriminan adalah membagi data menjadi data *training* dan *testing* dengan

proporsi 80:20 tiap kategori. Didapatkan data *training* sebanyak 147 observasi dan data *testing* sebanyak 39 observasi. Setelah melakukan pembagian data, kemudian dilakukan uji asumsi pada data *training* dan deteksi multikolinieritas. Asumsi yang akan diuji adalah normal multivariat dan homogenitas matriks varian-kovarian. Uji asumsi normal multivariat menggunakan statistik uji *Mardia-test* dan uji asumsi homogenitas matriks varian kovarian menggunakan statistik uji *Box's M*.

Tabel 4.3 Rangkuman Pengujian Asumsi Data *Training*

Asumsi	Statistik Uji	<i>P value</i>	Kesimpulan
Normal	<i>Skewness</i>	740,654	0
Multivariat	<i>Kurtosis</i>	6,344	0
Homogenitas Matriks Varian- Kovarian	<i>Box's M</i>	356,464	0
			Data tidak Normal Multivariat Matriks Varian- Kovarian tidak homogen

Tabel 4.3 menunjukkan nilai *p-value* pada *skewness* dan *kurtosis* kurang dari 0,05, sehingga disimpulkan bahwa asumsi normal multivariat tidak terpenuhi. Nilai *p-value* pada *Box's M* kurang dari 0,05, sehingga asumsi homogenitas matriks varian-kovarian tidak terpenuhi. Kemudian dilakukan transformasi data untuk memenuhi kedua asumsi tersebut.

Tabel 4.4 Transformasi Box-Cox Variabel

Variabel	<i>Lambda Box-Cox</i>
(x_1) Usia	1
(x_2) Sistolik	-0,5
(x_3) Diastolik	0
(x_4) Denyut	0
(x_5) Napas	-0,5
(x_6) Suhu	1
(x_7) pH	2
(x_8) Natrium	1
(x_9) Glukosa	0

Tabel 4.4 (lanjutan)

Variabel	<i>Lambda Box-Cox</i>
(x_{10}) Hematokrit	1
(x_{11}) PO ₂	-0,5
(x_{12}) BUN	-0,5

Tabel 4.4 menunjukkan *rounded lambda* setiap variabel independen yang kemudian dijadikan acuan dalam transformasi variabel. Setelah melakukan transformasi kemudian diperiksa kembali asumsi normal multivariat dan homogenitas matriks varian-kovarian pada data *training*.

Tabel 4.5 Rangkuman Pengujian Asumsi data *Training*

Asumsi	Statistik Uji		<i>P value</i>	Kesimpulan
Normal	<i>Skewness</i>	405,298	0,067	Data Normal
Multivariat	<i>Kurtosis</i>	1,354	0,176	Multivariat
Homogenitas Matriks Varian- Kovarian	<i>Box's M</i>	286,574	0	Matriks Varian- Kovarian tidak homogen

Tabel 4.5 menunjukkan hasil uji asumsi normal mutivariat dan homogenitas matriks varian-kovarian setelah data ditransformasi. Didapatkan hasil bahwa asumsi normal mutivariat terpenuhi karena nilai *p-value* pada *skewness* dan *kurtosis* lebih dari 0,05, namun asumsi homogenitas matriks varian-kovarian tidak terpenuhi karena nilai *p-value* sama dengan 0. Analisis fungsi diskriminan tetap *robust* walaupun asumsi homogenitas varian-kovarian tidak terpenuhi (Ghazali, 2011) dan untuk mengetahui kebaikan algoritma PSO, maka analisis diskriminan linier tetap dilakukan.

Sebelum melakukan analisis diskriminan perlu dilakukan pemeriksaan perbedaan rata-rata secara statistik. Pengujian perbedaan rata-rata menggunakan statistik uji *Wilk's Lambda*. Hipotesis yang digunakan adalah sebagai berikut.

$$H_0 : \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu$$

$H_1 : \mu_i \neq \mu_j, i \neq j$ (minimal ada sepasang kelompok yang berbeda)

Tabel 4.6 Statistik Uji Perbedaan Rata-rata antar Kelompok

	Wilks' Lambda	F	<i>P value</i>
Usia	0,713	19,223	0,000
Sistolik	0,966	1,701	0,170
Diastolik	0,983	0,827	0,481
Denyut	0,958	2,096	0,103
Napas	0,949	2,557	0,058
Suhu	0,988	0,556	0,645
pH	0,834	9,495	0,000
Natrium	0,963	1,847	0,141
Glukosa	0,990	0,466	0,707
Hematokrit	0,997	0,137	0,938
PO ₂	0,960	2,002	0,116
BUN	0,849	8,496	0,000

Tabel 4.6 menunjukkan hasil perhitungan statistik uji perbedaan rata-rata antar kelompok pada semua variabel yang diduga berpengaruh terhadap penentuan kelompok kelas risiko pasien pneumonia. Keputusan yang diambil adalah gagal tolak H_0 apabila *P value* kurang dari 0,05, sehingga dapat disimpulkan hanya pada variabel usia, analisis gas darah atau pH dan BUN yang memiliki perbedaan rata-rata antar kelompok. Variabel tekanan darah distolik, denyut nadi, frekuensi napas, natrium, dan PO₂ yang secara visual berdasarkan *boxplot* antar kelompok terlihat terdapat perbedaan rata-rata, ternyata setelah dilakukan uji secara statistik tidak menunjukkan adanya perbedaan yang signifikan. Kemudian dilakukan pemeriksaan korelasi antar variabel independen yang ditunjukkan pada Tabel 4.7 sebagai berikut.

Tabel 4.7 Korelasi Antar Variabel Independen

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11
x2	-0,151 0,067										
x3	-0,011 0,893	-0,693 0									
x4	-0,157 0,057	0,223 0,007	-0,067 0,417								
x5	0,168 0,041	-0,164 0,048	0,030 0,716	-0,436 0,000							
x6	0,063 0,446	0,049 0,555	-0,073 0,379	0,193 0,019	-0,078 0,348						
x7	-0,195 0,018	-0,040 0,638	0,040 0,627	0,013 0,880	-0,166 0,045	0,107 0,198					
x8	0,123 0,139	-0,166 0,044	0,090 0,277	-0,022 0,795	0,064 0,443	-0,086 0,299	-0,330 0,000				
x9	0,002 0,977	-0,074 0,373	0,119 0,151	-0,030 0,719	0,086 0,301	-0,000 0,996	0,010 0,901	-0,091 0,275			
x10	0,010 0,904	-0,078 0,347	0,140 0,091	-0,108 0,193	0,040 0,630	-0,115 0,164	-0,405 0,000	0,164 0,047	0,015 0,861		
x11	0,063 0,451	0,039 0,640	0,024 0,776	-0,001 0,995	-0,151 0,069	0,045 0,587	0,215 0,009	-0,108 0,191	0,101 0,226	0,101 0,223	
x12	-0,306 0,000	0,036 0,669	-0,017 0,833	0,090 0,278	0,018 0,827	0,070 0,396	0,129 0,120	-0,262 0,001	-0,090 0,279	-0,011 0,893	0,020 0,814

Tabel 4.7 menunjukkan hampir semua pasangan variabel independen tidak ada yang memiliki korelasi yang sangat tinggi. Korelasi yang sangat tinggi merupakan salah satu tanda kemungkinan adanya kasus multikolinieritas, sehingga dapat disimpulkan bahwa tidak ada kecenderungan adanya kasus multikolinieritas pada data rekam medis pasien pneumonia. Setelah melakukan pemeriksaan korelasi antara variabel individu langkah selanjutnya adalah melakukan seleksi variabel. Seleksi variabel dilakukan karena dalam menentukan kelas risiko seorang pasien banyak metode yang dapat dijadikan patokan. Setiap metode penentuan kelas risiko memiliki variabel yang berbeda-beda. Diharapkan seleksi variabel dapat variabel-variabel mana saja yang dapat mengelompokkan kelas risiko secara tepat dan efektif. Seleksi variabel pada penelitian ini dilakukan dengan menggunakan *syntax* pada lampiran 5 dan lampiran 6.

4.2.1 Analisis Diskriminan dengan *Forward Selection*

Metode seleksi menggunakan *forward selection* dilakukan dengan menghitung nilai *Wilks' A* parsial, F_{hitung} parsial dan *P value* untuk masing-masing variabel independen. Variabel independen yang akan dimasukkan dalam fungsi diskriminan adalah variabel dengan nilai F_{hitung} parsial terbesar atau nilai *P value* kurang dari 0,05.

Tabel 4.8 menunjukkan bahwa terdapat enam tahap dalam seleksi variabel *forward selection*, dalam setiap tahap akan terpilih satu variabel yang akan dimasukkan dalam persamaan fungsi diskriminan. Pada tahap ke-0, dapat dilihat bahwa variabel x_1 , x_7 , dan x_{12} memiliki *P value* kurang dari 0,05 namun variabel pertama memiliki nilai F_{hitung} terbesar. Setelah memasukkan variabel x_1 kedalam model, langkah selanjutnya adalah menghitung nilai *Wilks' A* pada variabel yang tersisa. Pada tahap 1, variabel yang memiliki nilai *P value* kurang dari 0,05 dan nilai F_{hitung} yang terbesar adalah x_7 . Tahap berhenti ketika semua variabel dengan

P value kurang dari 0,05 atau variabel yang signifikan masuk ke dalam model.

Tabel 4.8 Tahapan *Forward Selection* Data Pasien Pneumonia

Tahap	<i>Wilks' A</i>	<i>F</i> hitung	<i>P-value</i>	Variabel	
				Tersisa	Masuk
0	0,7126	19,2229	0	1	
	0,9656	1,7006	0,1696	2	
	0,9829	0,8272	0,4809	3	
	0,9579	2,0958	0,1034	4	
	0,9491	2,5565	0,0576	5	
	0,9885	0,5561	0,6449	6	
	0,8339	9,4946	0	7	1
	0,9627	1,8466	0,1414	8	
	0,9903	0,4655	0,7068	9	
	0,9971	0,1374	0,9375	10	
	0,9597	2,0017	0,1164	11	
	0,8487	8,4962	0	12	
1	0,9282	3,6611	0,014	2	
	0,9820	0,8685	0,4591	3	
	0,9168	4,2930	0,0062	4	
	0,8696	7,0983	0,0002	5	
	0,9913	0,4132	0,7437	6	
	0,8637	7,4726	0,0001	7	7
	0,9357	3,2554	0,0236	8	
	0,9868	0,6317	0,5957	9	
	0,9965	0,1684	0,9176	10	
	0,9620	1,8673	0,1378	11	
5	0,9166	4,3049	0,0061	12	
	⋮	⋮	⋮	⋮	⋮
	0,9621	1,8128	0,1477	2	
	0,9854	0,6798	0,5658	3	
	0,9581	2,0115	0,1152	4	
	0,9805	0,9154	0,4353	6	0
	0,9871	0,6024	0,6145	9	
	0,9609	1,8726	0,1371	10	
	0,9800	0,9379	0,4242	11	

Berdasarkan Tabel 4.8 variabel yang masuk dalam *forward selection* adalah x_1 , x_7 , x_5 , x_8 , dan x_{12} . Kemudian dari variabel yang terpilih dilakukan analisis diskriminan. Fungsi diskriminan yang terbentuk dari variabel usia, analisis gas darah atau pH, napas, natrium, dan BUN adalah sebagai berikut.

$$\hat{Z}_1 = -41,5957 - 0,0619x_1 + 30,896x_5 + 0,4820x_7 \\ + 0,0787 x_8 + 6,6129x_{12}$$

$$\hat{Z}_2 = 24,6805 - 0,0395x_1 - 6,0775x_5 - 0,5268x_7 \\ + 0,0594 x_8 - 1,7683x_{12}$$

$$\hat{Z}_3 = -15,2819 - 0,0180x_1 - 29,2081x_5 + 0,3378x_7 \\ + 0,0357 x_8 - 6,1920x_{12}$$

Dari fungsi diskriminan diatas kemudian dilakukan klasifikasi pada data *testing* menggunakan aturan dari persamaan (2.19). Hasil klasifikasi ditunjukkan pada Tabel 4.9 sebagai berikut.

Tabel 4.9 *Confusion Matrix* Klasifikasi Diskriminan *Forward* Data *Testing*

		<i>Classification Result</i>			
		K2	K3	K4	K5
<i>Actual</i>	K2	13	3	1	0
	K3	2	5	4	1
	K4	0	2	2	4
	K5	0	0	0	2

Tabel 4.9 menunjukan bahwa nilai *recall* (R) untuk setiap kategori didapatkan dari perhitungan nilai yang tepat diklasifikasikan dibagi dengan total observasi setiap kategori. Masing-masing nilai R sebagai berikut.

$$R_1 = \frac{13}{17} \times 100\% = 76,5\%; R_2 = 41,7\%; R_3 = 25\%; R_4 = 100\%$$

Kemudian dihitung nilai akurasi klasifikasi. Akurasi klasifikasi yang digunakan adalah *G-mean* dan *Area Under ROC Curve* (AUC). Perhitungan akurasi menggunakan kriteria *G-mean*

sangat sensitif apabila jumlah data *testing* pada salah satu kategori sangat kecil, maka untuk mengatasi permasalahan tersebut dilakukan pula perhitungan akurasi menggunakan kriteria AUC.

Pehitungan akurasi menggunakan *G-mean* adalah sebagai berikut.

$$GM = (0,76 \times 0,42 \times 0,25 \times 1)^{1/4} = 53,13\%$$

Perhitungan akurasi menggunakan kriteria AUC adalah sebagai berikut.

$$AUC = \frac{1}{4} (0,765 + 0,417 + 0,25 + 1) = 60,78\%$$

Maka didapatkan nilai akurasi menggunakan kriteria *G-mean* sebesar 53,13% dan akurasi menggunakan kriteria AUC sebesar 60,78%.

4.2.2 Analisis Diskriminan dengan *Backward Elimination*

Dalam *backward elimination*, langkah awal yang harus dilakukan adalah menghitung nilai nilai *Wilks' A* parsial, F_{hitung} parsial dan *P value* untuk masing-masing variabel independen. Kemudian mengeluarkan satu per satu variabel yang memiliki nilai F_{hitung} yang paling kecil atau memiliki nilai *P value* yang lebih besar dari 0,05. Pada tahap 0 terlihat bahwa x_2 , x_3 , x_4 , x_6 , x_9 , x_{10} , dan x_{11} memiliki nilai *P value* lebih dari 0,05 namun x_3 memiliki nilai F_{hitung} yang paling kecil. Langkah selanjutnya adalah menghilangkan variabel x_3 . Setelah menghapus variabel x_3 dari model, langkah selanjutnya adalah menghitung nilai *Wilks' A* parsial dari variabel yang tersisa. Pada tahap 1 terlihat bahwa variabel x_6 adalah variabel dengan nilai F_{hitung} terkecil dan *P value* lebih dari 0,05, maka variabel x_6 dikeluarkan dari model. Tahapan *backward elimination* berhenti apabila sudah tidak ada variabel yang memiliki nilai *P value* lebih besar dari 0,05. Ilustrasi tahapan seleksi *backward elimination* ditunjukkan pada tabel 4.10 sebagai berikut.

Tabel 4.10 Tahapan *Backward Elimination* Data Pasien Pneumonia

Tahap	Wilks 'A	F hitung	P-value	Variabel	
				Tersisa	Keluar
0	0,6953	19,2846	0	1	3
	0,9724	1,2486	0,2948	2	
	0,9912	0,3911	0,7596	3	
	0,9817	0,8219	0,484	4	
	0,8785	6,0869	0,0007	5	
	0,9892	0,4789	0,6975	6	
	0,7972	11,1898	0	7	
	0,8926	5,2969	0,0018	8	
	0,9852	0,6597	0,5782	9	
	0,9739	1,1791	0,3203	10	
	0,9823	0,7914	0,5008	11	
	0,8856	5,6864	0,0011	12	
1	0,6973	19,2485	0	1	6
	0,9714	1,307	0,2748	2	
	0,9806	0,8788	0,4539	4	
	0,8773	6,1984	0,0006	5	
	0,9890	0,4933	0,6875	6	
	0,7984	11,1958	0	7	
	0,8917	5,384	0,0016	8	
	0,9839	0,7255	0,5385	9	
	0,9765	1,0669	0,3655	10	
	0,9825	0,7912	0,5009	11	
	0,8847	5,7801	0,0010	12	
⋮	⋮	⋮	⋮	⋮	⋮
7	0,6989	20	0	1	0
	0,8200	10	0	5	
	0,7832	13	0	7	
	0,8692	7	0,0002	8	
	0,8845	6	0,0007	12	

Berdasarkan Tabel 4.10 variabel yang terpilih untuk masuk dalam model adalah x_1 , x_5 , x_7 , x_8 , dan x_{12} . Variabel yang terpilih untuk dimasukkan ke dalam fungsi diskriminan pada seleksi *backward elimination* sama dengan variabel terpilih hasil seleksi *forward selection*, maka fungsi diskriminan pada metode *backward elimination* sama seperti fungsi diskriminan pada analisis diskriminan *forward selection* (4.2.1) dengan akurasi yang sama, baik menggunakan kriteria *G-mean* yaitu sebesar 53,13% maupun akurasi menggunakan kriteria AUC yang bernilai 60,78%.

4.2.3 Analisis Diskriminan dengan *Stepwise Method*

Seleksi dengan *stepwise method* merupakan gabungan dari tahapan seleksi variabel menggunakan *forward selection* dan *backward elimination*. *Forward selection* digunakan untuk mendapatkan variabel yang akan dimasukkan ke dalam fungsi diskriminan dan *backward elimination* digunakan untuk mengevaluasi setiap penambahan variabel yang dimasukkan ke dalam fungsi diskriminan. Tahapan dalam *stepwise method* ditunjukkan dalam Tabel 4.11 sebagai berikut.

Tabel 4.11 Tahapan *Stepwise Method* Data Pasien Pneumonia

Tahap	Variabel Masuk	<i>F</i> hitung	<i>Wilks' Lambda</i>
1	x1	19,223	
2	x1	16,800	0,834
	x7	7,473	0,713
3	x1	22,926	0,757
	x7	10,246	0,620
	x5	9,855	0,615
4	x1	26,049	0,722
	x7	11,813	0,581
	x5	10,409	0,567
	x8	4,552	0,509

Tabel 4.11 (lanjutan)

Tahap	Variabel Masuk	F hitung	Wilks' Λ
5	x1	19,959	0,587
	x7	12,829	0,524
	x5	10,168	0,500
	x8	6,974	0,472
	x12	6,053	0,464

Tabel 4.11 menunjukkan bahwa variabel yang terpilih dalam fungsi diskriminan adalah x_1 , x_7 , x_5 , x_8 , dan x_{12} . Variabel yang terpilih dari seleksi variabel menggunakan *stepwise method* sama dengan hasil seleksi variabel baik menggunakan *forward selection* maupun *backward elimination*. Dari semua metode seleksi variabel tersebut terpilih lima variabel dari dua belas variabel awal. Variabel yang terpilih tersebut adalah Usia, Napas, pH, Natrium dan BUN. Akurasi klasifikasi pada analisis diskriminan dengan memasukkan lima variabel yang terpilih menggunakan kriteria *G-mean* sebesar 53,13% dan akurasi menggunakan kriteria AUC sebesar 60,78%.

4.2.4 Analisis Diskriminan - PSO


Selain melakukan seleksi variabel dengan menggunakan metode *Forward Selection*, *Backward Elimination*, dan *Stepwise Method*, algoritma *Particle Swarm Optimization* (PSO) juga digunakan untuk seleksi variabel hingga didapatkan nilai *Fitness* terbaik. Seleksi variabel algoritma PSO dilakukan menggunakan *syntax* pada lampiran 7. Nilai *fitness* dalam algoritma PSO merupakan kesalahan klasifikasi yang didapat dari perhitungan $1 - G\text{-mean}$ dan $1 - AUC$. Berikut ini merupakan ilustrasi langkah-langkah dalam seleksi variabel dengan algoritma PSO dengan kriteria *fitness* $1 - G\text{-mean}$.

1. Menentukan nilai parameter-parameter dalam algoritma PSO.
Nilai koefisien akselerasi c_1 dan c_2 adalah 2. Bobot inertia

yang digunakan sebesar 0,9. Bobot inersia ini dipilih berdasarkan penelitian yang dilakukan Lin dan Chen (2009) yang menyebutkan bahwa bobot tersebut memberikan hasil optimum yang lebih baik dibanding bobot inertia 0,7, 0,8 atau 1. Jumlah populasi partikel dalam penelitian ini adalah 50 partikel dengan jumlah iterasi sebesar 100 kali.

2. Menentukan posisi awal partikel dan kecepatan awal partikel. Pada posisi awal partikel tidak bergerak dalam kecepatan tertentu, sehingga kecepatan awal semua partikel adalah 0. Posisi awal atau inisialisai digunakan untuk menentukan variabel independen mana yang terpilih untuk dimasukkan di dalam model, di mana kode 1 berarti variabel tersebut masuk ke dalam model dan kode 0 berarti variabel tersebut tidak dimasukkan ke dalam model. posisi awal dihitung dengan merandom nilai 0 hingga 1. Apabila angka random bernilai lebih dari atau sama dengan 0,5 maka variabel tersebut terpilih dan disimbolkan dengan kode 1 dan apabila bernilai kurang dari 0,5 maka variabel dihapuskan dalam model. Ilustrasi pada salah satu partikel adalah sebagai berikut.

x_1	x_2	x_3	...	x_{10}	x_{11}	x_{12}
0,3280	0,2198	0,0663	...	0,0784	0,6301	0,2153



x_1	x_2	x_3	...	x_{10}	x_{11}	x_{12}
0	0	0	...	0	1	0

Gambar 4.16 Representasi Inisialisasi Posisi Partikel PSO pada Iterasi Awal

Satu partikel terdiri atas 12 variabel solusi. Setiap partikel akan menghasilkan variabel-variabel terpilih yang kemudian akan dihitung nilai kesalahan klasifikasi tiap partikel.

Tabel 4.12 Ilustrasi Populasi Awal Partikel PSO

Partikel	x_1	x_2	x_3	x_4	...	x_9	x_{10}	x_{11}	x_{12}
1	0	0	0	0	...	1	0	1	0
2	1	0	0	0	...	0	0	1	1
3	0	0	0	0	...	1	1	0	0
\vdots	\vdots	\vdots	\vdots	\vdots		\vdots	\vdots	\vdots	\vdots
49	1	1	1	0	...	1	0	1	0
50	1	0	1	1	...	0	0	0	0

Tabel 4.12 menunjukkan inisialisasi posisi pada setiap partikel pada iterasi awal. Selanjutnya untuk mengevaluasi masing-masing partikel maka dihitung nilai *fitness* yang merupakan hasil perhitungan kesalahan klasifikasi data *testing* pada analisis diskriminan dari variabel yang terpilih tiap partikel. *Fitness* atau kesalahan klasifikasi dihitung dengan menggunakan rumus $1 - G\text{-Mean}$. Hasil evaluasi nilai *fitness* pada setiap partikel ditunjukkan oleh Tabel 4.13 sebagai berikut.

Tabel 4.13 Ilustrasi Nilai *Fitness* Setiap Partikel Populasi Awal PSO

Partikel	x_1	x_2	x_3	x_4	...	x_9	x_{10}	x_{11}	x_{12}	<i>Fitness</i>
1	0	0	0	0	...	1	0	1	0	0,5765
2	1	0	0	0	...	0	0	1	1	0,5621
3	0	0	0	0	...	1	1	0	0	0,5181
\vdots	\vdots	\vdots	\vdots	\vdots		\vdots	\vdots	\vdots	\vdots	\vdots
49	1	1	1	0	...	1	0	1	0	0,4905
50	1	0	1	1	...	0	0	0	0	0,5696

3. Dari semua partikel pada iterasi awal dicatat nilai *local best* dan *global best*. Pada iterasi awal, semua posisi awal merupakan *local best*. *Global best* adalah nilai terbaik dari semua partikel, maka pada iterasi awal ini nilai *global best* adalah partikel dengan nilai *fitness* 0,3846. Partikel yang

menjadi *global best* pada iterasi awal ditunjukkan dalam ilustrasi sebagai berikut.

x_1	x_2	x_3	...	x_{10}	x_{11}	x_{12}
0,4201	0,2364	0,4765	...	0,3171	0,1027	0,7547

x_1	x_2	x_3	...	x_{10}	x_{11}	x_{12}
0	0	0	...	0	0	1

Gambar 4.17 Representasi Posisi *Global Best* pada Iterasi Awal

4. Langkah selanjutnya adalah menghitung kecepatan partikel untuk bergerak pada iterasi selanjutnya. Partikel bergerak berdasarkan informasi nilai *local best* dan *global best*. Kecepatan partikel dihitung menggunakan persamaan (2.22). Ilustrasi perhitungan kecepatan partikel adalah sebagai berikut.


$$\begin{aligned}
 v_1^2 = 0,9 \cdot \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} + 2 \cdot \begin{bmatrix} 0,62 \\ 0,51 \\ \vdots \\ 0,32 \\ 0,51 \end{bmatrix}^T \cdot \left(\begin{bmatrix} 0,24 \\ 0,22 \\ \vdots \\ 0,63 \\ 0,22 \end{bmatrix} - \begin{bmatrix} 0,24 \\ 0,22 \\ \vdots \\ 0,63 \\ 0,22 \end{bmatrix} \right) \\
 + 2 \cdot \begin{bmatrix} 0,24 \\ 0,17 \\ \vdots \\ 0,93 \\ 0,42 \end{bmatrix}^T \cdot \left(\begin{bmatrix} 0,42 \\ 0,24 \\ \vdots \\ 0,10 \\ 0,76 \end{bmatrix} - \begin{bmatrix} 0,24 \\ 0,22 \\ \vdots \\ 0,63 \\ 0,22 \end{bmatrix} \right) = \begin{bmatrix} 0,09 \\ 0,01 \\ \vdots \\ -0,99 \\ 0,45 \end{bmatrix}
 \end{aligned}$$

Kecepatan tiap partikel dibatasi v_{dmax} sebesar 4 dan v_{dmin} sebesar -4. Kemudian partikel akan berpindah pada posisi baru berdasarkan kecepatan yang telah dihitung. Perpindahan partikel dihitung dengan persamaan (2.21).


$$x_1^2 = \begin{bmatrix} 0,24 \\ 0,22 \\ \vdots \\ 0,63 \\ 0,22 \end{bmatrix} + \begin{bmatrix} 0,09 \\ 0,01 \\ \vdots \\ -0,99 \\ 0,45 \end{bmatrix} = \begin{bmatrix} 0,33 \\ 0,23 \\ \vdots \\ -0,36 \\ 0,67 \end{bmatrix}$$

Posisi partikel dibatasi antara 0 hingga 1. Apabila nilai posisi partikel lebih dari satu maka akan diubah menjadi 1, begitu pula apabila posisi bernilai kurang dari 0 maka akan diubah menjadi 0. Sehingga pada iterasi kedua ini posisi partikel adalah sebagai berikut.

x_1	x_2	x_3	...	x_{10}	x_{11}	x_{12}
0,51	0,25	0,54	...	0,76	-0,89	1,21



x_1	x_2	x_3	...	x_{10}	x_{11}	x_{12}
0,51	0,25	0,54	...	0,76	0,00	1,00



x_1	x_2	x_3	...	x_{10}	x_{11}	x_{12}
1	0	1	...	1	0	1

Gambar 4.18 Representasi Posisi Partikel pada Iterasi Kedua

- Setiap partikel akan bergerak sebanyak 100 kali iterasi. Setiap kali bergerak setiap partikel akan menyimpan *fitness* terbaik mereka yang kemudian disebut sebagai *local best* dan setiap partikel akan mengirimkan informasi *local best* ke partikel lain sehingga dapat ditentukan nilai terbaik dari keseluruhan partikel yang disebut sebagai *global best*. *Local best* dan *global best* selalu di *update* setiap iterasi dan menjadi penentuan dalam perhitungan kecepatan untuk menuju posisi selanjutnya. Rangkuman hasil terbaik dalam setiap iterasi ditunjukkan dalam Tabel 4.14 sebagai berikut.

Tabel 4.14 Seleksi Variabel dengan Algoritma PSO dengan *Fitness* G-mean

Iterasi ke-	x1	x2	x3	x4	...	x9	x10	x11	x12	<i>Fitness</i>
1	0	0	0	1	...	0	0	0	1	0,3846
2	1	0	1	1	...	1	0	1	1	0,3731

Tabel 4.14 (lanjutan)

Iterasi ke-	x1	x2	x3	x4	...	x9	x10	x11	x12	<i>Fitness</i>
3	1	0	1	0	...	1	1	0	0	0,3264
⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮	⋮	⋮
99	1	0	1	0	...	1	1	0	0	0,3264
100	1	0	1	0	...	1	1	0	0	0,3264

Tabel 4.14 menunjukkan bahwa nilai *Fitness* konvergen mulai iterasi ke-3 dengan *Fitness* terbaik adalah 0,326. Variabel yang terpilih berdasarkan fitness terbaik tersebut adalah Usia, Diastolik, Napas, Analisis Gas Darah atau pH, Glukosa, dan Hematokrit. Fungsi diskriminan dengan variabel hasil seleksi menggunakan algoritma PSO dengan kriteria *fitness* *G-mean* adalah sebagai berikut.

$$\begin{aligned}\hat{Z}_1 &= -32,7611 - 0,0673x_1 + 0,7634x_3 + 34,1135x_5 \\ &\quad + 0,5008x_7 - 0,5411x_9 + 0,0446x_{10} \\ \hat{Z}_2 &= 46,5111 - 0,0455x_1 - 1,1649x_3 - 1,0517x_5 \\ &\quad - 0,6651x_7 - 0,1786x_9 - 0,0342x_{10} \\ \hat{Z}_3 &= -1,6866 + 0,0091x_1 + 2,3434x_3 + 35,0631x_5 \\ &\quad - 0,2264x_7 + 0,059x_9 - 1,6866x_{10}\end{aligned}$$

Setelah mendapatkan fungsi diskriminan dari variabel yang terpilih berdasarkan algoritma *particle swarm optimization* (PSO), kemudian dilakukan klasifikasi pada data *testing* menggunakan aturan dari persamaan (2.19). Hasil klasifikasi pada data *testing* ditunjukkan pada Tabel 4.15 sebagai berikut.

Tabel 4.15 *Confusion Matrix* Klasifikasi ADL-PSO Data *Testing*

		<i>Classification Result</i>			
		K2	K3	K4	K5
<i>Actual</i>	K2	14	2	1	0
	K3	1	8	3	0
	K4	1	2	3	2
	K5	0	0	0	2

Tabel 4.15 menunjukkan bahwa nilai *recall* (R) untuk setiap kategori didapatkan dari perhitungan nilai yang tepat diklasifikasikan dibagi dengan total observasi setiap kategori. Masing-masing nilai R sebagai berikut.

$$R_1 = \frac{14}{17} \times 100\% = 82,4\%; R_2 = 66,7\%; R_3 = 37,5\%; R_4 = 100\%$$

Maka didapatkan nilai GM sebagai berikut.

$$GM = (0,824 \times 0,667 \times 0,375 \times 1)^{1/4} = 67,36\%$$

Maka didapatkan akurasi klasifikasi berdasarkan kriteria *G-mean* sebesar 67,36% atau kesalahan klasifikasi sebesar $100\% - 67,36\% = 32,64\%$.

Langkah yang sama juga dilakukan pada seleksi variabel algoritma PSO dengan kriteria *fitness* $1 - AUC$. Rangkuman hasil terbaik dalam setiap iterasi algoritma PSO dengan kriteria AUC ditunjukkan dalam Tabel 4.16 sebagai berikut.

Tabel 4.16 Seleksi Variabel dengan Algoritma PSO dengan *Fitness* AUC

Iterasi ke-	x1	x2	x3	x4	...	x9	x10	x11	x12	<i>Fitness</i>
1	1	0	1	0	...	1	1	0	0	0,3045
2	1	0	1	0	...	1	1	0	0	0,3045
3	1	0	1	0	...	1	1	0	0	0,3045
⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮	⋮	⋮
99	1	0	1	0	...	1	1	0	0	0,2837
100	1	0	1	0	...	1	1	0	0	0,2837

Hasil seleksi variabel dengan Algoritma PSO menggunakan kriteria *fitness* AUC menunjukkan bahwa nilai *fitness* konvergen mulai iterasi ke-19 dengan *fitness* terbaik adalah nilai *fitness* terkecil yaitu 0,2837. Variabel yang terpilih berdasarkan *fitness* AUC terbaik sama dengan variabel terpilih berdasarkan *fitness* *G-mean*, variabel tersebut adalah Usia, Diastolik, Napas, Analisa Gas Darah atau pH, Glukosa, dan Hematokrit. Fungsi diskriminan dengan variabel hasil seleksi menggunakan algoritma PSO dengan kriteria *fitness* AUC sama dengan fungsi diskriminan dengan

variabel hasil seleksi menggunakan algoritma PSO dengan kriteria fitness G-mean

Ringkasan hasil analisis diskriminan dengan seleksi variabel dengan menggunakan metode *forward selection*, *backward elimination*, *stepwise method* dan algoritma *particle swarm optimization* (PSO) ditunjukkan dalam Tabel 4.17 sebagai berikut.

Tabel 4.17 Hasil Seleksi Variabel

Metode Seleksi Variabel	Variabel Terpilih	Akurasi	
		G-mean	AUC
<i>Forward Selection</i>	$x_1, x_5, x_7, x_8, x_{12}$	0,5313	0,6078
<i>Backward Elimination</i>	$x_1, x_5, x_7, x_8, x_{12}$	0,5313	0,6078
<i>Stepwise Method</i>	$x_1, x_5, x_7, x_8, x_{12}$	0,5313	0,6078
<i>Algoritma PSO</i>	$x_1, x_3, x_5, x_7, x_9, x_{10}$	0,6736	0,7163

Variabel terpilih dari hasil seleksi variabel yang telah dilakukan, masing-masing metode seleksi *variabel forward selection*, *backward elimination*, *stepwise method*, dan Algoritma PSO dilakukan analisis diskriminan untuk mendapatkan fungsi diskriminan multi-class. Hasil estimasi parameter dan nilai kesalahan klasifikasi dari keempat metode seleksi variabel pada klasifikasi disajikan pada Tabel 4.18 sebagai berikut.

Tabel 4.18 Rangkuman Estimasi Parameter Fungsi Diskriminan

Metode		1	2	3	Akurasi	
					G-mean	AUC
<i>Forward, Backward, dan Stepwise</i>	$w_1 (x_1)$	-0,062	-0,040	-0,018	0,5313	0,6078
	$w_2 (x_5)$	30,896	-6,078	-29,208		
	$w_3 (x_7)$	0,482	-0,527	0,338		
	$w_4 (x_8)$	0,079	0,059	0,036		
	$w_5 (x_{12})$	6,613	-1,768	-6,192		
	α	-41,596	24,680	-15,272		

Tabel 4.18 (lanjutan)

Metode		1	2	3	Akurasi	
					G-mean	AUC
Algoritma PSO	$w_1 (x_1)$	-0,067	-0,046	0,009	0,6736	0,7163
	$w_2 (x_3)$	0,763	-1,165	2,343		
	$w_3 (x_5)$	34,114	-1,052	35,063		
	$w_4 (x_7)$	0,501	-0,665	-0,226		
	$w_5 (x_9)$	-0,541	-0,179	0,006		
	$w_6 (x_{10})$	0,045	-0,034	-0,084		
	α	-32,761	46,511	-1,687		
Semua Variabel	$w_1 (x_1)$	0,063	-0,033	-0,022	0,5450	0,6140
	$w_2 (x_2)$	32,677	35,504	-78,162		
	$w_3 (x_3)$	0,627	-0,070	-3,988		
	$w_4 (x_4)$	1,029	1,689	0,774		
	$w_5 (x_5)$	-26,876	3,814	-19,461		
	$w_6 (x_6)$	0,090	0,203	0,431		
	$w_7 (x_7)$	-0,544	-0,334	0,219		
	$w_8 (x_8)$	-0,067	0,064	0,018		
	$w_9 (x_9)$	0,336	-0,008	-0,094		
	$w_{10} (x_{10})$	-0,036	-0,006	0,063		
	$w_{11} (x_{11})$	-0,250	-16,075	8,542		
	$w_{12} (x_{12})$	-6,444	-1,066	-5,940		
	α	28,333	-5,408	-6,067		

Tabel 4.18 menunjukkan bahwa metode seleksi variabel menggunakan *forward selection*, *backward elimination*, dan *stepwise method* memberikan nilai akurasi klasifikasi G-mean sebesar 53,13% dan akurasi klasifikasi AUC sebesar 60,78%. Sedangkan pada analisis diskriminan dengan melibatkan semua variabel menghasilkan nilai akurasi klasifikasi yang sedikit lebih baik dibanding metode *forward*, *backward*, dan *stepwise* dengan akurasi G-mean sebesar 54,50% dan akurasi AUC sebesar 61,40%. Nilai akurasi klasifikasi terbaik yaitu metode seleksi variabel menggunakan algoritma PSO yang memberikan nilai akurasi klasifikasi G-mean sebesar 67,36% dan akurasi AUC sebesar

71,63%. Hal ini menunjukkan bahwa hanya dengan mengikutsertakan enam variabel Usia, Diastolik, Napas, Analisa Gas Darah atau pH, Glukosa, dan Hematokrit sudah mampu mengklasifikasikan kelas resiko pasien pneumonia dengan akurasi yang lebih baik dibanding mengikutsertakan semua variabel.

4.3 Klasifikasi Naïve Bayes Pasien Pneumonia

Sebelum melakukan klasifikasi menggunakan Naïve Bayes, langkah pertama yang harus dilakukan adalah membagi data menjadi data *training* dan *testing* dengan proporsi 80:20 tiap kategori pada variabel dependen. Didapatkan data *training* sejumlah 147 observasi dan data *testing* sebanyak 39 observasi. Kemudian dilakukan klasifikasi Naïve Bayes, baik untuk semua variabel maupun variabel yang terpilih sebelumnya menggunakan *forward selection*, *backward elimination*, *stepwise method*, dan algoritma PSO.

Pada klasifikasi Naïve Bayes menggunakan semua variabel sebelum dilakukan transformasi, setelah data telah dibagi menjadi data *training* dan data *testing* langkah selanjutnya adalah mencari peluang bersyarat pada tiap kategori kelas risiko untuk setiap variabel independen. Untuk mencari peluang bersyarat tiap kategori digunakan rumus (2.28) karena semua variabel independen adalah variabel kontinu. Diperlukan nilai rata-rata dan standar deviasi pada data *training* untuk menghitung semua peluang pada tiap kategori. Rata-rata dan standar deviasi dapat dilihat pada Tabel 4.19 sebagai berikut.

Tabel 4.19 Rata-rata dan Standar Deviasi Tiap Kategori Kelas Risiko Data *Training*

	Rata-rata				Standar Deviasi			
	K2	K3	K4	K5	K2	K3	K4	K5
x_1	45,84	59,54	62,26	70,00	13,04	13,29	10,87	12,51
x_2	121,36	122,72	115,29	115,83	19,35	18,90	26,80	27,28
x_3	74,83	75,78	72,65	70,00	10,56	12,64	14,84	12,65

Tabel 4.19 (lanjutan)

	Rata-rata				Standar Deviasi			
	K2	K3	K4	K5	K2	K3	K4	K5
x_4	103,56	102,11	109,58	113,50	15,70	17,01	16,39	12,45
x_5	26,75	28,30	29,55	33,33	5,50	7,23	7,80	5,61
x_6	37,05	37,05	37,15	37,48	0,79	0,84	1,01	0,92
x_7	7,45	7,44	7,35	7,35	0,07	0,08	0,13	0,10
x_8	137,58	134,68	137,52	137,02	4,82	6,74	9,26	10,61
x_9	151,42	168,22	167,10	171,00	86,15	94,68	92,48	99,44
x_{10}	38,28	37,57	37,98	38,78	5,80	6,50	7,14	8,66
x_{11}	111,49	91,15	129,54	91,05	56,18	32,79	71,66	25,79
x_{12}	14,51	19,16	27,26	42,00	8,58	14,33	17,47	25,12

Nilai rata-rata dan standar deviasi dari data *training* tiap kategori kelas risiko dalam Tabel 4.18 digunakan untuk menentukan peluang tiap kategori pada data *testing* menggunakan persamaan (2.28). Data testing ditunjukkan dalam Lampiran 2. Perhitungan peluang data *testing* pada tiap kategori kelas risiko dilakukan sebagai berikut.

Tabel 4.20 menunjukkan peluang parsial tiap kategori kelas risiko pada data *testing* pertama.

Tabel 4.20 Peluang Parsial pada Tiap Kategori Data *Testing* Pertama

	x_{1i}	$P(x_{1i} K2)$	$P(x_{1i} K3)$	$P(x_{1i} K4)$	$P(x_{1i} K5)$
x_1	54	0,02516	0,02752	0,02751	0,01408
x_2	100	0,01121	0,01025	0,01265	0,01236
x_3	60	0,01410	0,01448	0,01870	0,02308
x_4	100	0,02477	0,02328	0,02052	0,01781
x_5	24	0,06399	0,04622	0,03972	0,01782
x_6	36,5	0,39595	0,38297	0,31991	0,24506
x_7	7,37	3,21919	3,34279	3,12001	3,90676
x_8	143,1	0,04291	0,02713	0,03594	0,03192
x_9	97	0,00379	0,00318	0,00324	0,00304

Tabel 4.20 (lanjutan)

	x_{1i}	$P(x_{1i} K2)$	$P(x_{1i} K3)$	$P(x_{1i} K4)$	$P(x_{1i} K5)$
x_{10}	41,4	0,05954	0,05159	0,04981	0,04404
x_{11}	112	0,00710	0,00994	0,00540	0,01112
x_{12}	12	0,04457	0,02458	0,01560	0,00778

Setelah menemukan peluang parsial tiap variabel pada setiap kategori kelas risiko, langkah selanjutnya adalah menghitung peluang posterior yang kemudian digunakan untuk menentukan klasifikasi kelompok kelas risiko pada data *testing*.

$$P(X_1, X_2, \dots, X_{12}|K2) = P(x_{11}|K2). P(x_{12}|K2) \dots P(x_{12}|K2) \\ = 0,025 \times 0,011 \times \dots \times 0,044 = 2,45 \times 10^{-17}$$

$$P(X_1, X_2, \dots, X_{12}|K3) = P(x_{11}|K3). P(x_{12}|K3) \dots P(x_{12}|K3) \\ = 0,028 \times 0,010 \times \dots \times 0,024 = 6,12 \times 10^{-18}$$

$$P(X_1, X_2, \dots, X_{12}|K4) = P(x_{11}|K4). P(x_{12}|K4) \dots P(x_{12}|K4) \\ = 0,028 \times 0,013 \times \dots \times 0,016 = 2,59 \times 10^{-18}$$

$$P(X_1, X_2, \dots, X_{12}|K5) = P(x_{11}|K5). P(x_{12}|K5) \dots P(x_{12}|K5) \\ = 0,014 \times 0,012 \times \dots \times 0,008 = 4,52 \times 10^{-18}$$

Hasil perhitungan peluang posterior pada data *testing* pertama menunjukkan bahwa peluang terbesar adalah kategori kelas risiko II maka data tersebut akan diklasifikasikan ke dalam kelas risiko II. Perhitungan tersebut dilakukan untuk semua data *testing*. Hasil dari klasifikasi ditunjukkan pada Tabel 4.21 sebagai berikut.

Tabel 4.21 *Confusion Matriks* Klasifikasi Naive Bayes Data *Testing*

		<i>Classification Result</i>			
		K2	K3	K4	K5
<i>Actual</i>	K2	11	2	3	1
	K3	3	5	4	0
	K4	2	3	2	1
	K5	0	0	1	1

Berdasarkan Tabel 4.20 terlihat bahwa nilai *recall* (R) untuk setiap kategori didapatkan dari perhitungan nilai yang tepat diklasifikasikan dibagi dengan total observasi setiap kategori. Masing-masing nilai R sebagai berikut.

$$R_1 = \frac{11}{17} \times 100\% = 64,7\%; R_2 = 41,7\%; R_3 = 25\%; R_4 = 50\%$$

Kemudian dihitung nilai akurasi klasifikasi. Akurasi klasifikasi yang digunakan adalah *G-mean* dan *Area Under ROC Curve* (AUC). Perhitungan akurasi menggunakan kriteria *G-mean* akan sangat sensitif apabila jumlah data *testing* pada salah satu kategori sangat kecil, maka untuk mengatasi permasalahan tersebut dilakukan pula perhitungan akurasi menggunakan kriteria AUC.

Pehitungan akurasi menggunakan *G-mean* adalah sebagai berikut.

$$GM = (0,65 \times 0,42 \times 0,25 \times 1)^{1/4} = 42,85\%$$

Perhitungan akurasi menggunakan kriteria AUC adalah sebagai berikut.

$$AUC = \frac{1}{4} (0,65 + 0,42 + 0,25 + 1) = 45,34\%$$

Maka di dapatkan nilai akurasi menggunakan kriteria *G-mean* sebesar 53,13% dan akurasi menggunakan kriteria AUC sebesar 60,78%.

Kemudian dilakukan perhitungan yang sama untuk data dengan variabel yang sudah ditransformasi dan variabel yang terseleksi baik seleksi dengan metode *forward*, *backward*, dan *stepwise* serta seleksi dengan PSO. Rangkuman hasil klasifikasi menggunakan Naïve Bayes ditunjukkan dalam Tabel 4.22 sebagai berikut.

Tabel 4.22 Hasil Klasifikasi Naïve Bayes

Variabel Terlibat	Jenis Data	Akurasi	
		G-mean	AUC
Semua Variabel	Asli	0,4285	0,4534
	Transformasi	0,4484	0,4743

Tabel 4.22 (lanjutan)

Variabel Terlibat	Jenis Data	Akurasi	
		G-mean	AUC
<i>Forward, Backward, dan Stepwise</i> ($x_1, x_5, x_7, x_8, x_{12}$)	Asli	0,0000	0,4559
	Transformasi	0,5271	0,5411
Seleksi PSO ($x_1, x_3, x_5, x_7, x_9, x_{10}$)	Asli	0,0000	0,4914
	Transformasi	0,0000	0,4559

Tabel 4.21 menunjukkan hasil akurasi klasifikasi Naïve Bayes baik menggunakan kriteria *G-mean* maupun AUC. Pada klasifikasi Naïve Bayes menggunakan variabel yang terpilih dari hasil seleksi *Forward Selection, Backward Elimination, dan Stepwise Methode*, untuk data asli dan klasifikasi menggunakan variabel seleksi PSO memiliki kesalahan klasifikasi sebesar 100% hal ini disebabkan hanay terdapat 2 data *testing* pada kategori kelas risiko 5 dan setelah diklasifikasikan tidak ada yang masuk kelas risiko 5. Akurasi terbaik berdasarkan kriteria *G-mean* adalah klasifikasi Naïve Bayes dengan mengikutsertakan hasil seleksi variabel *stepwise selection, backward elimination dan stepwise method* dengan nilai akurasi sebesar 52,71%. Sama seperti kriteria akurasi *G-mean*, klasifikasi Naïve Bayes terbaik dengan menggunakan kriteria akurasi AUC adalah klasifikasi dengan mengikutsertakan variabel hasil seleksi variabel *stepwise selection, backward elimination dan stepwise method*. Variabel tersebut adalah Usia, Napas, pH, Natrium dan BUN.

4.4 Perbandingan Hasil Klasifikasi Analisis Diskriminan dan Naïve Bayes

Setelah melakukan klasifikasi pada data pasien pneumonia menggunakan analisis diskriminan dengan seleksi variabel *forward selection, backward elimination dan stepwise methode, hybrid* analisis diskriminan dengan algoritma PSO dan *Naïve Bayes Classification*, langkah selanjutnya adalah membandingkan

hasil akurasi dari metode tersebut. Ringkasan hasil klasifikasi dan variabel yang terlibat ditunjukkan dalam Tabel 4.23 sebagai berikut.

Tabel 4.23 Akurasi Klasifikasi Tiap Metode

Metode Klasifikasi	Variabel	Akurasi	
		G-mean	AUC
Analisis Diskriminan	Semua Variabel	0,5450	0,6140
	<i>Forward, Backward dan Stepwise</i> ($x_1, x_5, x_7, x_8, x_{12}$)	0,5313	0,6078
	Algoritma PSO ($x_1, x_3, x_5, x_7, x_9, x_{10}$)	0,6736	0,7163
Naive Bayes dengan Data Asli	Semua Variabel	0,4285	0,4534
	<i>Forward, Backward dan Stepwise</i> ($x_1, x_5, x_7, x_8, x_{12}$)	0,0000	0,4559
	Algoritma PSO ($x_1, x_3, x_5, x_7, x_9, x_{10}$)	0,0000	0,4914
Naive Bayes dengan Data Transformasi	Semua Variabel	0,4484	0,4743
	<i>Forward, Backward dan Stepwise</i> ($x_1, x_5, x_7, x_8, x_{12}$)	0,5271	0,5441
	Algoritma PSO ($x_1, x_3, x_5, x_7, x_9, x_{10}$)	0,0000	0,4559

Tabel 4.23 menunjukkan akurasi setiap metode klasifikasi. Metode klasifikasi yang menghasilkan akurasi terbaik berdasarkan kriteria akurasi G-mean adalah *hybrid* analisis diskriminan dengan algoritma *Particle Swarm Optimization* (PSO) dengan nilai akurasi sebesar 67,36%. Sama seperti kriteria akurasi G-mean, metode klasifikasi dengan kriteria akurasi AUC adalah *hybrid* analisis diskriminan dengan algoritma *Particle Swarm Optimization* (PSO) dengan akurasi sebesar 71,63%.

(Halaman ini sengaja dikosongkan)

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dipaparkan pada BAB IV maka dapat diambil kesimpulan sebagai berikut.

1. Variabel yang terpilih untuk dimasukan fungsi diskriminan menggunakan metode *forward selection*, *backward elimination* dan *stepwise method* adalah usia, frekuensi napas, analisis gas darah arteri atau pH, kadar natrium darah dan BUN yang menghasilkan akurasi *G-mean* sebesar 53,13% dan akurasi AUC sebesar 60,78%. Sedangkan variabel yang terpilih dari seleksi menggunakan algoritma PSO baik menggunakan kriteria *fitness* $1 - G\text{-mean}$ maupun $1 - AUC$ adalah usia, diastolik, napas, analisi gas darah atau ph, glukosa, dan hematokrit dengan akurasi *G-mean* sebesar 67,36% dan akurasi AUC sebesar 71,63%. Sedangkan analisis diskriminan dengan melibatkan semua variabel menghasilkan akurasi *G-mean* sebesar 54,5% dan akurasi AUC sebesar 61,4%.
2. Proses klasifikasi menggunakan metode Naïve Bayes terbaik adalah dengan melibatkan variabel usia, frekuensi napas, analisis gas darah arteri atau pH, kadar natrium darah dan BUN yang merupakan hasil seleksi variabel menggunakan *forward selection*, *backward elimination* dan *stepwise method* setelah dilakukan transformasi dengan akurasi *G-mean* sebesar 52,72% dan akurasi AUC sebesar 54,11%.

3. Metode klasifikasi yang menghasilkan nilai akurasi terbesar baik menggunakan kriteria akurasi *G-mean* maupun kriteria akurasi AUC adalah *hybrid* analisis diskriminan dengan *Particle Swarm Optimization* (PSO).

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, saran yang dapat diberikan oleh peneliti adalah sebagai berikut.

1. Lebih teliti dalam melakukan pencatatan data pasien pneumonia.
2. Dalam melakukan klasifikasi terhadap kelas risiko pasien pneumonia disarankan menggunakan metode yang dapat mengikutsertakan semua variabel yang terlibat, baik variabel kontinyu maupun variabel kategorik dan metode dengan asumsi yang sesuai.
3. Mengumpulkan lebih banyak data pasien pneumonia, sehingga jumlah data dalam setiap kategori diharapkan dapat *balance*.

DAFTAR PUSTAKA

- Alva, J. A., & Estrada, E. G. (2009). A Generalization of Shapiro–Wilk's Test for Multivariate Normality. *Communications in Statistics - Theory and Method*, 38:11, 1870-1883. doi:10.1080/03610920802474465.
- American Lung Association. (2015). *Trend in Pneumonia and Influenza Morbidity and Mortality*. America: American Lung Association Research and Helath Education Division.
- Anonim. (2015). *Infodatin Situasi Penyebab Kanker* . Jakarta Selatan: Kemertrian Kesehatan RI.
- Ashari, A., Paryudi, I., & Tjoa, A. (2013). Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool, Vol 4. *International Journal of Advanced Computer Science and Applications*, 33-39. Retrieved from IJACSA thesai database.
- Aujesky, D., & Fine, M. (2008). *The Pneumonia Severity Index: A Decade After the Initial Derivation and Validation*. America: Infectious Diseases Society of America.
- Bai, Q. (2010). *Analysis of Particle Swarm Optimization Algorithm*. China: Inner Mongolia University for Nationalities.
- Bekkar, M., Djemaa, H., & Alitouch, T. (2013). Evaluation Measures for Models Assesment over Imbalanced Data Sets, Vol. 3. *Journal of Information Engineering and Application* , 27-38. Retrieved from International Knowledge Sharing Platform.
- Engelbrecht, A. (2007). *Computational Intelligence - An Intriduction 2nd Edition*. Wiley.
- Fatimah , E., Sukarsa, I., & Susilawati, M. (2015). *Pemodelan Resiko Penyakit Pneumonia pada Balita di Provinsi*

- Jawa Timur dengan Pendekatan Geographically Wighted Logistic Regression*. Bali: Universitas Udayana.
- Ghozali, I. (2011). *Aplikasi Analisis Multivariate dengan Program IBM SPSS 19 (Edisi Kelima)*. Semarang: Universitas Diponegoro.
- Gorunescu, F. (2011). *Data Mining Concept, Models and Techniques*. Australia: Springer-Verlag Berlin Heidelberg.
- Hair, J., Black, W., & Babin, B. (2010). *Multivariate Data Analysis 7th Edition*. New Jersey: Pearson Prentice Hall.
- Hockenberry, M., & Wilson, D. (2009). *Essential of Pediatric Nursing*. Mosby: St. Louis Missouri.
- Hocking, R. R. (2003). *Methods and Application of Linear Models 2nd Edition*. New Jersey: John Wiley & Sons.
- Johnson, R., & Wichern, D. (2007). *Applied Multivariate Statistical Analysis*. New Jersey: Pearson Education.
- Kemenkes RI. (2010). *Pneumonia Balita. Jendela Epidemiologi*, 1.
- Kennedy, J., & Eberhart, R. (1995). *Particle Swarm Optimization*. Wasington: Purdue School of Engineering and Technology
- Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014) *MVN: An R Package for Assessing Multivariate Normality*. The R Journal.6(2):151-162. Web site: <https://journal.r-project.org/archive/2014-2/korkmaz-goksuluk-zararsiz.pdf>
- Kurnianto, I.P, Irhamah. (2016). *Seleksi Variabel Dan Estimasi Parameter Menggunakan Metode Hybrid Analisis Diskriminan Multi-Class Dan Algoritma Genetika Untuk Klasifikasi*. Surabaya: Institut Teknologi Sepuluh Nopember.

- Kurniawan, L., & Israr, Y. (2009). *Pneumonia Pada Dewasa*. Riau: Universitas Riau.
- Lin, S., & Chen, S. (2009). PSOLDA: A Particle Swarm Optimization Approach for Enhancing Classification Accuracy Rate of Linear Discriminant Analysis. *Applied Soft Computing* 9, 1008-1015. doi:10.1016/j.asoc.2009.01.001
- Marino. (2014). *Perbandingan Analisis Diskriminan Linier, Diskriminan Linier Robust dan Regresi Logistik Biner Studi Kasus Pada Penjurusan Bidang IPA / IPS Tingkat SMA Negeri 1 Bangarejo Banyuwangi*. Jember: Universitas Jember.
- Misnadiarly. (2008). *Penyakit Infeksi Saluran Napas Pneumonia pada Balita, Orang Dewasa, Usia Lanjut*. Jakarta: Pustaka Obor Populer.
- Noviana, I., Wulandari, S., & Purhadi. (2013). *Pemodelan Resiko Penyakit Pneumonia pada Balita di Jawa Timur Menggunakan Regresi Logistik Biner Stratifikasi*. Surabaya: Intitut Teknologi Sepuluh Nopember.
- PDPI. (2003). *Pneumonia Komuniti Pedoman Diagnosis dan Penatalaksanaan di Indonesia*. Perhimpunan Dokter Paru Indonesia.
- Santoso. (2010). *Statistik Multivariat Konsep dan Aplikasi dengan SPSS*. Jakarta: PT Elex Media Komputindo.
- Schoene, T. (2011). *Step-Optimized Particle Swarm Optimization*. Saskatoon: University of Saskatchewan.
- Setyoningrum, R. (2006). *Pneumonia*. In *Continuing Education Ilmu Kesehatan Anak XXXVI*. Surabaya: SMF Ilmu Kesehatan Anak FK Unair RSUD Dr. Soetomo.
- Shi, Y., & Eberhart, R. (1998). A Modified Particle Swarm Optimizaer. *IEEE Word Congress on Coputational Intelligence*, 1998:69-73. doi: 10.1109/ICEC.1998.699146

- Sivakumar, S., & Chandrasekar, C. (2014). *Modified Based Feature Selection for Classification of Lung CT Images*. India: Periyar University.
- Subadana, I., & Purniti, N. (2010). *Faktor-Faktor yang Berhubungan dengan Pneumonia pada Anak*. Denpasar: Universitas Udayana.
- Sun, Y., Kamel, M., & Wang, Y. (2006). Boosting for Learning Multiple Classes with Imbalanced Class Distribution. *International Conference on Data Mining (ICDM'06)*. Retrieved from Electrical and Computer Engineering Duke University Database.
- Tambayong, J. (1999). *Patofisiologi untuk Keperawatan*. Jakarta: Penerbit Buku Kedokteran EGC.
- Velez, J., Correa, J., & Ramos, F. 2015. A New Approach to the Box-Cox transformation. *Frontier in Applied Mathematic and Statistics*, 1-10. doi:10.3389/fams.2015.00012.
- Wicaksana, P. D. (2015). *Perbandingan Algoritma K-Nearest Neighbors dan Naive Bayes untuk Data Winconsin Diagnosis Breast Cancer*. Yogyakarta: Universitas Sanata Dharma.

Lampiran 1 Data *Training* Pasien Pneumonia

Usia	Sistolik	Diastolik	...	HCT	PO ₂	BUN	Kelas
24	100	70	...	30,2	147	18	2
47	100	60	...	41,8	100	11	2
58	125	86	...	44,3	54,8	14	2
50	100	70	...	40,4	56,4	15	2
55	120	80	...	41,3	101	10	2
44	110	80	...	42,5	135	21	2
⋮	⋮	⋮	...	⋮	⋮	⋮	⋮
48	130	80	...	46,1	77	14	2
35	80	40	...	32,3	86,8	23	3
67	130	100	...	36,7	69,3	10	3
46	110	70	...	41,7	43,3	10	3
71	110	70	...	46,2	80	24	3
59	130	70	...	28,9	114,5	19	3
63	150	80	...	39,4	76,4	48	3
⋮	⋮	⋮	...	⋮	⋮	⋮	⋮
43	130	80	...	37,8	36	13	3
70	100	70	...	48,7	68	31	4
72	120	60	...	53,1	86	13	4
55	130	90	...	44,7	54	55	4
64	115	80	...	39,7	133	40	4
56	100	80	...	42,3	56	22	4
89	138	82	...	41	68	58	4
⋮	⋮	⋮	...	⋮	⋮	⋮	⋮
75	80	50	...	45,5	66	21	4
73	150	80	...	44,7	59	15	5
73	115	70	...	51,9	126,7	88	5
73	140	80	...	27,4	106,6	44	5
85	90	50	...	35,4	88	28	5
47	80	60	...	39,6	65	45	5
69	120	80	...	33,7	101	32	5

Lampiran 2 Data *Testing* Pasien Pneumonia

Usia	Sistolik	Diastolik	...	HCT	PO ₂	BUN	Kelas
54	100	60	...	41,4	112	12	2
42	110	80	...	37,9	64	12	2
39	100	70	...	33,9	106	77	2
48	110	80	...	45,5	73	6	2
64	110	80	...	44,5	96	18	2
47	95	55	...	29,3	56,5	4	2
⋮	⋮	⋮	...	⋮	⋮	⋮	⋮
65	185	100	...	39,7	83	9	2
43	130	98	...	42,7	71	17	3
64	130	88	...	51,5	220,4	22	3
66	120	60	...	45,3	161	29	3
61	110	80	...	42,5	69	10	3
60	110	70	...	44,2	89	11	3
72	120	80	...	38,1	100	21	3
⋮	⋮	⋮	...	⋮	⋮	⋮	⋮
55	110	70	...	30,3	90	18,8	3
78	110	60	...	37,4	79	33	4
58	110	60	...	36,4	275	18	4
72	170	80	...	41,7	67	22	4
61	130	80	...	36,8	54,2	13	4
61	100	70	...	44,2	189	16	4
⋮	⋮	⋮	...	⋮	⋮	⋮	⋮
78	140	100	...	24,3	59	17	4
74	100	60	...	28,7	105,7	51	5
77	140	90	...	46,6	135	29	5

Lampiran 3 *Syntax* Estimasi Parameter Fungsi Diskriminan

```

diskriminan=function(datatraining,datatesting)
{
library(MASS)
library(caret)
data=datatraining
nn=ncol(data)
data2=datatesting
grouptesting=data2[,nn]
training=data[,1:(nn-1)]
testing=data2[,1:(nn-1)]
if (nn<3){
training = as.matrix(training)
testing = as.matrix(testing)
}
else
{
training = training
testing = testing
}
p=ncol(training)
n=nrow(training)
nts=nrow(testing)
g=4

meantrain=matrix(0,p)
for(i in 1:(nn-1))
{
meantrain[i]=sum(training[,i])/n
}
grouptraining=data[,nn]
z=lda(training,grouptraining)
kkk=predict(z,training)$class
acu=confusionMatrix(grouptraining,kkk)
acu1=as.matrix(acu)
dacc=diag(acu1)

```

Lampiran 3 (lanjutan)

```

tepat=1
for (i in 1:g)
{
tepat=tepat*(dacc[i]/sum(acu1[i,]))
}
salahtr=1-(tepat^(1/g))
beta=as.matrix(z$scaling)
konstanta=-(t(meantrain)%*%beta)
rownames(konstanta)=c("kons")
estimasi=rbind(beta,konstanta)

a=as.matrix(estimasi)
a=t(a)
g=4
d=2
grouptest=matrix(0,g)

for (i in 1:g)
{
y1=matrix(0,nrow=nts)
for (j in 1:nts)
{
if(grouptesting[j]==d){
y1[j]=1
}
}
d=d+1
grouptest[i]=sum(y1)
}
gmean=matrix(0,g,p)
xbar=matrix(0,g,p)
for (k in 1:g)
{
for (j in 1:p)
{

```

Lampiran 3 (lanjutan)

```

gmean[k,j]=mean(training[grouptraining==(k+1),j]) #rata-rata
perkelompok tiap var prediktor
}
}
b=0
y=matrix(0,nts,(g-1))
for (i in 1:(g-1))
{
for (j in 1:nts)
{
c=0
for (k in 1:p)
{
c=c+(testing[j,k]%%a[i,k])
}
y[j,i]=c+a[i,nn]
}
}
z=matrix(0,g,g-1)
for (i in 1:(g-1))
{
for (j in 1:g)
{
c=0
for (k in 1:p))
{
c=c+(gmean[j,k]*a[i,k]);
}
z[j,i]=c+a[i,nn];
}
}
x=matrix(0,nts,g)

for (i in 1:g)
{

```

Lampiran 3 (lanjutan)

```

for (j in 1:nts)
{
c=0
for (k in 1:(g-1))
{
c=c+(y[j,k]-z[i,k])^2
}
x[j,i]=c
}
}
p=0
for (i in 1:nts)
{
s=min(x[i,])
w=1
v=x[i,w]

while (v>s)
{
w=w+1
v=x[i,w]
}
p[i]=w+1
}
acc=confusionMatrix(grouptesting,p)
acc2=as.matrix(acc)
dacc=diag(acc2)
tepat=1
for (i in 1:g)
{
tepat=tepat*(dacc[i]/grouptest[i])
}
salah=1-(tepat^(1/g))
list(estimasiparameter=estimasi,kesalahantraining=salahtr,kesalahan
ntesting=salah,Confusion_Matriks=acc)
}

```

Lampiran 4 Syntax Pembagian *Training* dan *Testing*

```

crossval=function(data)
{

nn=ncol(data)
data=data[order(data[,nn]),]
datay=data[,nn]
n=nrow(data)
k=5
subset=seq(1,n,1)
nl=matrix(0,4)
ns=matrix(0,4)
sn=0
for(l in 2:5)
{
koding=data[subset[datay==l],]
nl[l]=nrow(koding)
sn=sn+nrow(koding)
ns[l]=sn
}
a=1
fold=0
for (i in 2:5)
{
folds=cvFolds(nl[i],K=k)
fold[a:ns[i]]=folds$which
a=a+nl[i]
}
datax=cbind(data,fold)

datatraining=datax[subset[datax[, (nn+1)]!=1],]
datetesting=datax[subset[datax[, (nn+1)]=1],]
list(training=datatraining[, (1:nn)],testing=datetesting[, (1:nn)])
}

```

Lampiran 5 Syntax Seleksi Variabel *Forward Selection*

```
function forwardbaru( data )
nn = length(data(1,:));
data = sortrows(data,nn);
x = data(:,1:(nn-1));
y1 = data(:,nn);
y = y1-1;
[n,p] = size(x);
vari = 1:1:p;
[y,groups,glevels] = grp2idx(y); %gindex menunjukkan
kategori Y, groups me-list groupnya apa saja, glevels
menunnnjukan level groupnya berapa saja
g = length(groups);
n1 = hist(y,1:g); %menunjukkan jumlah observasi tiap
group
for k = 1:g
    gmeans(k,:) = mean(x(y==k,:),1); %rata2 per
kelompok tiap variabel independen
end
Q1 = x-gmeans(y,:); %awal mula untuk menghitung W
xbar = repmat(mean(x),n,1);
Total = x-xbar;
SW = Q1'*Q1;
SWN = SW/(n-g);
ST = Total'*Total;
SB = ST-SW;
SBN = SB/(g-1);
WLM = zeros(p,3);
for i = 1:p
    WLM(i,1) = SW(i,i)/ST(i,i);
    WLM(i,2) = SBN(i,i)/SWN(i,i);
    WLM(i,3) = 1-fcdf(WLM(i,2),g-1,n-g);
end
WLM(:,4) = vari;
j=0;
disp(j);
WLM

%mengecek variabel yang signifikan
for i = 1:p
    if WLM(i,3)<0.05
        WLM(i,2) = WLM(i,2);
    else
        WLM(i,2) = 0;
    end
end
end
```

Lampiran 5 (lanjutan)

```

Fmak = max(WLM(:,2));
if Fmak == 0;
    'No Variable Enter'
else
    %mengambil X yang masuk dalam model
    x1 = x;
    x2 = x;
    sumj = 0;
    for i = 1:p
        if WLM(i,2)>=Fmak;
            sumj = sumj+1;
            xx = (x1(:,i));
            x2(:,i) = 0;
            variabel = vari(i);
            var2 = vari;
            var2(i) = [];
        end
    end
    v1 = 1;
    for i = 1:p
        if x2(:,i) == 0;
            v1 = v1;
        else
            x(:,v1) = x2(:,i);
            v1 = v1+1;
        end
    end
    data1 = x(:,1:(p-sumj)); %untuk menyimpan data yang
    belum masuk dalam model
    xx1 = xx; %untuk menyimpan variabel terpilih
    [s1,s2] = size(data1);
    for i = 1:s2
        [a1,a2] = size(xx1);
        [b1,b2] = size(data1);
        gmeans1 = NaN(g,b2);
        gmeans2 = NaN(g,a2);
        for k = 1:g
            gmeans1(k,:) = mean(data1(y==k,:),1);
            gmeans2(k,:) = mean(xx1(y==k,:),1);
        end
        %Data Variabel yang belum masuk model
        R1 = data1-gmeans1(y,:); %awal mula untuk
menghitung W
        Rat1 = repmat(mean(data1),n,1);
        Tot1 = data1-Rat1;
        SW1 = R1'*R1;
    end
end

```

Lampiran 5 (lanjutan)

```

ST1 = Tot1'*Tot1;
    %Data Variabel masuk model
    R2 = xx1-gmeans2(y,:); %awal mula untuk menghitung W
    Rat2 = repmat(mean(xx1),a1,1);
    Tot2 = xx1-Rat2;
    SW2 = R2'*R2;
    ST2 = Tot2'*Tot2;
    Wilks2 = det(SW2)/det(ST2);
    WLMP = zeros(b2,4);
    for i = 1:b2
        CWP = [R1(:,i) R2];
        CTP = [Tot1(:,i) Tot2];
        WP = CWP'*CWP;
        TP = CTP'*CTP;
        WilP = det(WP)/det(TP);
        WLMP(i,1) = WilP/Wilks2;
        WLMP(i,2) = ((1-WLMP(i,1))/WLMP(i,1))*((n-g-
a2)/(g-1));
        WLMP(i,3) = 1-fcdf(WLMP(i,2),g-1,n-g-a2);
        WLMP(i,4) = WilP;
        WLMP(i,5) = var2(i);
    end
    j = j+1;
    disp(j)
    WLMP
    %memilih F terbesar
    for i = 1:b2
        if WLMP(i,3)<0.05;
            WLMP(i,2) = WLMP(i,2);
        else
            WLMP(i,2) = 0;
        end
    end
    Fmak1 = max(WLMP(:,2));
    if Fmak1 == 0;
        xx1 = xx1;
        datal = datal;
    else
        for i = 1:b2
            if WLMP(i,2)>=Fmak1;
                xx1=[xx1 datal(:,i)];
                datal(:,i) = [];
                variabel = [variabel var2(i)];
                var2(i) = [];
            end
        end
    end
end %untuk for q
end
variabel
end

```


Lampiran 6 Syntax Seleksi Variabel *Backward Elimination*

```
function backwardbaru( data )
nn = length(data(1,:));
data = sortrows(data,nn);
x = data(:,1:(nn-1));
y1 = data(:,nn);
y = y1-1;
variabel = [];
[n,p] = size(x);
vari = 1:1:p;
var2 = vari;
for i = 1:p
    disp(i-1)
    [n,d] =size(x);
    [y,groups,glevels] = grp2idx(y); %gindex menunjukkan
    kategori Y, groups me-list groupnya apa saja, glevel
    menunjukkan level groupnya berapa saja
    g = length(groups);
    nl = hist(y,1:g); %menunjukkan jumlah observasi tiap
    group
    gmeans = NaN(g,d);
    for k = 1:g
        gmeans(k,:) = mean(x(y==k,:),1); %rata2 per
        kelompok tiap variabel prediktor
    end

    Q = x-gmeans(y,:); %awal mula untuk menghitung W
    xbar = repmat(mean(x),n,1);
    Total = x-xbar;
    SW = Q'*Q;
    ST = Total'*Total;
    Wilks = det(SW)/det(ST);
    WLP = zeros(d,4);
    WLP = zeros(d,4);
    for i = 1:d
        Q1 = Q;
        Q1(:,i) = [];
        Total1 = Total;
        Total1(:,i) = [];
        SW1 = Q1'*Q1;
        ST1 = Total1'*Total1;
        Wilks1 = det(SW1)/det(ST1);
        WLP(i,1) = Wilks/Wilks1;
        WLP(i,2) = ((1-WLP(i,1))/WLP(i,1))*((n-g-
        d+1)/(g-1));
        WLP(i,3) = 1-fcdf(WLP(i,2),g-1,n-g-d+1);
        WLP(i,4) = Wilks1;
```

Lampiran 6 (lanjutan)

```

end
WLP(:,4) = var2;
WLP

%Memilih F terkecil
for i = 1:d
    if WLP(i,3)>=0.05
        WLP(i,2)=WLP(i,2);
    else
        WLP(i,2)=NaN;
    end
end
Fmin = min(WLP(:,2));
if Fmin == NaN
    x = x;
else
    x1 = x;
    sumj=0
    for i = 1:d
        if WLP(i,2)<=Fmin
            sumj = sumj+1;
            x1(:,i) = 0;
            variabel = vari(i);
            var2(i) = [];
        end
    end
    v1 =1;
    for i = 1:d
        if x1(:,i) == 0;
            v1 = v1;
        else
            x(:,v1) = x1(:,i);
            v1 = v1+1;
        end
    end
    x = x(:,1:(d-sumj));
end
end %milik for i=1:d
variabel

end

```

Lampiran 7 Syntax Algoritma PSO

```

#pso parameters
pso = list(
  nP = 50,          #population size
  min = rep(0,12), #lower bound for initial parameters
  max = rep(1,12), #upper bound for initial parameters
  maxV = 4,         #lower bound for velocity
  minV = -4,        #upper bound for velocity
  c1 = 2.0,
  c2 = 2.0,
  iner = 0.9,       #inertia
  initV = 1.00,     #intertial velocity
  d = 12            #number of parameters
)

OF = function(mP,datatraining,datatesting)
{
  nn = ncol(datatraining)
  dataa = datatraining[,1:(nn-1)]
  dataa2 = datatesting[,1:(nn-1)]
  grtr = datatraining[,nn]
  grts = datatesting[,nn]
  tr = dataa
  ts = dataa2
  bTr = nrow(tr)
  bTs = nrow(ts)
  nB = array(0,pso$nP)
  salah = array (0,pso$nP)

  for (k in 1:pso$nP)
  {
    for (i in 1:pso$d)
    {
      if(mP[i,k]>0.5)
      {
        nB[k]=nB[k]+1
      }
    }
    tra = array(0,c(bTr,nB[k]))
    tes = array(0,c(bTs,nB[k]))
    j = 1
  }
}

```

Lampiran 7 (lanjutan)

```

    for (i in 1:psod)
    {
      if (mP[i,k]>0.5)
      {
        tra[,j]=tr[,i]
        tes[,j]=ts[,i]
        j = j+1
      }
    }
  datatr = cbind(tra,grtr)
  datats = cbind(tes,grts)
  salah[k] = diskriminan(datatr,datats)$kesalahantesting
}
salah
}

PSO = function(pso,datatraining,datatesting,nG)
{
  #set up initial population and velocity
  mP = pso$min+diag(pso$max-pso$min)%*%mRU(pso$d,pso$nP)
  mV = array(0,c(pso$d,pso$nP))

  #make sure minimum one var to be choosed
  mD = array(0,c(pso$d,pso$nP))
  for (k in 1:pso$nP)
  {
    for (i in 1:psod)
    {
      if (mP[i,k]<0.5)
      {
        mD[i,k]=0
      }
      else
      {
        mD[i,k]=1
      }
    }
  }
}

```

Lampiran 7 (lanjutan)

```

while(sum(mD[,k])==0)
{
  mP[,k]=runif(5,min=0,max=1)
  for (i in 1:pso$d)
  {
    if (mP[i,k]<0.5)
    {
      mD[i,k]=0
    }
    else
    {
      mD[i,k]=1
    }
  }
}
#evaluate initial population
vF = OF(mP,datatraining,datatesting)

#set up best solution
mPbest = mP                                     #matrix of 'personally
best' solution
vFbest = vF                                     #vector of OF of best
solution
sGbest = min(vFbest)                             #scalar: best OF-value
sgbest = which.min(vFbest)[1]                   #scalar: best solution (counter)
Best = array(0,nG)
varbest = array(0,c(pso$d,nG))
#start generation
for (g in 1:nG){

  # update population
  mDV = pso$c1*mRU(pso$d,pso$nP)*(mPbest-
mP)+pso$c2*mRU(pso$d,pso$nP)*(mPbest[,sgbest]-mP)
  mV = pso$iner*mV+mDV

  #apply lower bound and upper bound limit for velocity
  for (k in 1:pso$nP)
  {
    for (i in 1:pso$d)

```

Lampiran 7 (lanjutan)

```

        {
            if (mV[i,k]>pso$maxV)
            {
                mP[i,k]=pso$maxV
            }
        }
        if (mP[i,k]<pso$minV)
        {
            mP[i,k]=pso$minV
        }
    }
}

mP = mP+mV

#apply lower bound and upper bound limit for position
for (k in 1:pso$nP)
{
    for (i in 1:pso$d)
    {
        if (mP[i,k]>1)
        {
            mP[i,k]=1
        }
    }
    if (mP[i,k]<0)
    {
        mP[i,k]=0
    }
}

#evaluate update population
vF = OF(mP,datatraining,datatesting)
vF[!(is.finite(vF))] = 10000 #a large value for NAs etc.
#find improvements
logik = vF < vFbest #improved solution
mPbest[,logik] = mP[,logik]
vFbest[logik] = vF[logik]
#find best solution
if (min(vF) < sGbest){
    sGbest = min(vF)
}

```

Lampiran 7 (lanjutan)

```
                                sgbest = which.min(vF)[1]
                                }
    Best[g] = sGbest
    varbest[,g] = mPbest[,sgbest]
    }           #end generation

    #return best solution
    return(list(OFvalue = Best, variabel = varbest))
}
```

Lampiran 8 Sertifikat Kode Etik Penelitian di RSUD Dr. Soetomo

F.LITB.003




KOMITE ETIK PENELITIAN KESEHATAN
RSUD Dr. SOETOMO SURABAYA
KETERANGAN KELAIKAN ETIK
(" ETHICAL CLEARANCE ")

104 / Panke.KKE/ II / 2017

KOMITE ETIK RSUD Dr. SOETOMO SURABAYA TELAH MEMPELAJARI SECARA SEKSAMA RANCANGAN PENELITIAN YANG DIUSULKAN, MAKA DENGAN INI MENYATAKAN BAHWA PENELITIAN DENGAN JUDUL :

" Klasifikasi Pasien Penderita Pneumonia Menggunakan Metode Analisis Diskriminan Linier,
*Hybrid Analisis Diskriminan Linier-Particle Swarm Optimization (ADL-PSO) dan
NAIVE BAYES Classification "*

PENELITI UTAMA : Neni Alya Firdausanti

PENELITI LAIN : 1. Dr. Irhamah, S.Si., M.Si
2. Dr. Laksmi Wulandari, dr., Sp.P (K), FCCP

UNIT / LEMBAGA / TEMPAT PENELITIAN : RSUD Dr. Soetomo Surabaya

DINYATAKAN LAIK ETIK

SURABAYA, 24 FEB 2017

KEKETUA



((Dr. Elizeus Hanindito, dr., Sp.An, KIC, KAP))
NIP. 19511007 197903 1 002